PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	)	
	:	Examiner: Manjunath N. Rao, Ph.D.
SUNG	)	
	:	Group Art Unit: 1652
Application No.: 09/990,874	)	
	:	Confirmation No. 2196
Filed: November 21, 2001	)	
	:	
For: XYLANASES WITH ENHANCED THERMOPHILICITY AND ALKALOPHILICITY	)	
	:	

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION UNDER 37 C.F.R. §1.132

I, Theresa White hereby declare as follows:

1) That I obtained my Ph.D. Biological in Chemistry, from Massachusetts Institute of Technology, Feb. 1991, and that I have over 15 years of experience in conducting and directing research in the areas of Biochemistry, Cell and Molecular biology, and Microbiology including Functional high-level expression of proteins in fungal, bacterial, yeast, and plant systems, Biochemical and structural studies on hydrolase and redox enzymes, Optimization of enzyme properties by directed and random mutagenesis, Optimization of fungal strains and fermentation processes for enzyme production.

2) That I am a Senior Research Scientist, Iogen Corporation - Ottawa, Ontario, Canada from March 1994 – present. Prior to this position I was a Post-Doctoral Research Fellow,

Loeb Medical Research Institute - Ottawa, Ontario, Canada,from October 1993 - March 1994, and a Post-doctoral Fellow and Research Scientist, Agriculture Canada, Ottawa, Ontario, Canada from December 1990 - October 1993.

3) That I have reviewed and understand the specification and the prosecution history of the 09/990,874 application.

4) That this Declaration is submitted in order to clarify for the record what one of ordinary skill in the art would understood from the prior art at the time the present invention was made.

5) That glycosyl hydrolases are enzymes that catalyze hydrolysis of glycosidic linkages between covalently linked sugar moieties. Traditionally, enzymes are classified by the Enzyme Commission based upon the chemistry they catalyze (hydrolysis, oxidation, ammonium transfer, etc). In the case of glycosyl hydrolases, much effort has gone into the development of a systematic nomenclature that identifies the enzymes by their substrate preference(s) and by their catalytic mechanism and structural features. Starting in 1991, work conducted primarily by Bernard Henrissat showed that glycosyl hydrolase enzymes can be classified into Families in which members not only share a common three-dimensional structure, but also regions of conserved amino acid homology and common catalytic mechanisms (Davies and Henrissat, 1995, Structure 3: 853-859). Xylanases (EC3.2.1.8) were among the first group of glycosyl hydrolases classified based on this system and are found almost exclusively in Family 10 and Family 11. At present, there are over 70 Glycosyl Hydrolase families and this system of nomenclature has been adopted internationally for use in publications and sequence databases. As a result, one can know the three-dimensional structure and mechanism of a protein simply by knowing the Glycosyl Hydrolase family to which it belongs.

6) That proteins are classified as Family 11 xylanases if (a) they exhibit the ability to hydrolyze internal beta-1, 4 glycosidic bonds between adjacent xylose residues in the main chain of the xylan polymer with retention of configuration at the anomeric carbon and (b)

they exhibit the primary and secondary structural signatures associated with Family 11 xylanase. Prior to the filing date of the application referred to in paragraph 3 above, dozens of amino acid sequences for Family 11 xylanases had been determined (including those in Figure 1 in the current application). In addition, about a dozen three-dimensional structures of Family 11 xylanases had been determined by NMR and x-ray crystallography and their structures published and co-ordinates deposited in the protein data bank (see attached Figure 1). At the present time, there are numerous Family 11 xylanase structures in the Protein Data Bank. Attached Figure 1 shows the 3-dimensional structures of 5 fungal and three bacterial Family 11 xylanases from PDB files.

7) That the high conservation of sequence and structural homology that defines Family 11 xylanases continues to be supported in the literature. In 2002, Sapag et al. (2002, Journal of Biotechnology 95: 109-131) conducted amino acid sequence alignments of 82 Family 11 xylanases (most of which were available prior to the filing date of the application referred to in paragraph 3 above) ranging in length from 173 to 220 amino acids and identified highly conserved signature sequences in beta strands B5, B6 and B8 as well as in the alpha helix.

8) That in 2003, Hakulinen et al. (2003, Eur. J. Biochem. 270: 1399-1412) conducted pairwise structural comparisons of ten Family 11 xylanases exhibiting 31-97% identity in amino acid sequence using molecular co-ordinates of the main chain C-alpha atoms and showed that the root-mean-square deviation (rmsd) ranged from 0.6 to 1.4 Å for a given pair of structures. This level of deviation is within the typical resolution of most X-ray crystal structures indicating that natural deviation of three-dimensional structure of Family 11 xylanases is very small.

9) That all Family 11 xylanases contain two conserved glutamate residues at positions 86 and 177 (see Figure 1 of the application referred to in paragraph 3 above; based on *Trichoderma reesei* xylanase II (TrX II, or Tr2) amino acid numbering), which are located on beta-strands B4 and B5 (Torronen, et al., 1994, EMBO J. 13: 2493-2501).

10) That assays that specifically detect xylanase activity, such as the enzyme-dependent release of reducing sugars from xylan substrates as is disclosed in the present application, were widely available at the time of filing of the application referred to in paragraph 3 above.

11) That the functional expression of several Family 11 xylanases from bacterial and fungal sources had been demonstrated in *E. coli* and *Saccharomyces cerevisiae* (for example, Yang et al., 1989, Appl. Environ. Microbiol. 55: 1192-1195; LaGrange et al., 1996, Appl. Environ. Microbiol. 62: 1036-1044).

12) That one of ordinary skill in the art at the time of the filing of the application referred in paragraph 3 above, could develop a thermophilic, thermostable or alkalophilic Family 11 xylanase by

- using degenerate primers for the highly conserved signature sequences of Family 11 xylanases to clone the gene from a microbial source and functionally express the enzyme in an expression host such as *E. coli*;
- determine the sequence of the cloned Family 11 xylanase and conduct a BLAST alignment against other known Family 11 xylanase sequences to identify regions of high and low homology;
- use standard molecular biology methods to generate site-specific amino acid substitutions or replacement of several codons with homologous codons from other Family 11 xylanase genes; and
- express the modified xylanases from *E. coli* and assess their thermostability, thermophilicity and/or alkalophilicity using known xylanase assays.

13) That the thermostability, thermophilicity or alkalophilicity of a Family 11 xylanase can also be achieved by Directed Evolution, first developed in the late 1990's. Directed Evolution requires (a) functional expression of the xylanase gene in a host microbe that can be transformed with high efficiency, such as the *E. coli* expression systems for xylanases

presented above; (b) a method for randomly mutating the Family 11 xylanase gene with a controlled mutation frequency, such as by error-prone PCR or DNA "shuffling"; and (c) a highly selective high throughput screening assay that can identify active variants with targeted properties.

14) That the techniques required for Directed Evolution were available prior to the filing date as demonstrated by Chen et al. (2001, Can. J. Micro. 47: 1088-1094), who used Directed Evolution to produce an alkalophilic variant of the Family 11 xylanase from *Neocallimastic patriciarum*. In this study, xylan-containing agar was overlaid on top of *E. coli* colonies expressing variants of the *N. patriciarum* xylanase to detect those clones expressing active xylanases. Thus, one of skill can readily identify xylanase variants that retain activity from a large population of variants using a simple agar plate assay. Only those clones expressing active xylanases are picked and grown in liquid culture to produce xylanase variants that will be tested for improved thermostability, alkalophilicity or thermophilicity.

15) That one of ordinary skill in the art could develop a thermophilic, thermostable or alkalophilic Family 11 xylanase by

- using degenerate primers for the highly conserved signature sequences of Family 11 xylanases to clone the gene from a microbial source and functionally express the enzyme in an expression host such as *E. coli*;
- generate random mutagenesis libraries by error-prone PCR and clone the library of variants with high efficiency into an expression host such as *E. coli;*
- overlay the *E. coli* colonies with xylan-containing agar, incubate the plate and identify those colonies producing clearing zones, indicative of xylanase activity;
- pick and grow the clones producing active xylanase in liquid culture; freeze a portion of the culture for DNA isolation;

- prepare a cell lysate to release the xylanase variant produced by each culture and then conduct a standard xylanase assay at elevated temperature (thermophilicity), elevated pH (alkalophilicity) or under optimal pH and temperature after a brief exposure to elevate temperature (thermostability); and

- isolate the DNA from the save portion of the liquid cultures for those clones expressing improved variants, perform DNA sequencing and identify the favorable mutation.

16) That there is not only sufficient information concerning the conservation of three-dimensional structures and sequences among Family 11 xylanases, but there is also a large body of information concerning their functional expression in common cloning hosts (e.g., *E. coli*) and a variety of assay formats for detecting their activity.

17) That, determining whether or not a given mutation is compatible with the mutations at positions 116, 118, 144 and 161, as presently claimed in the application referred to in paragrach 3 above, can be determined with ease by one of skill in the art by measuring the thermophilicity, alkalophilicity and/or thermostability after introducing such mutation(s).

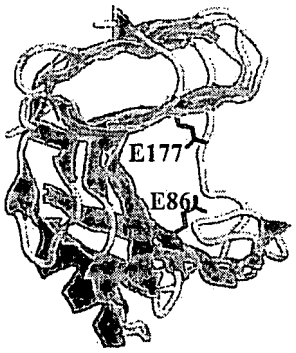
18) That I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine and imprisonment, or both, Under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon

EXECUTED at Ottawa, Canada this 20 day of June, 2007.

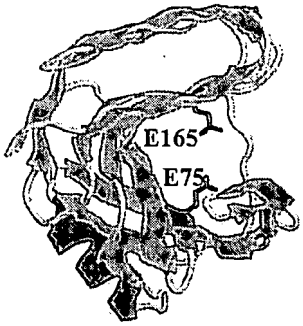
By _____

Figure 1 (attached):

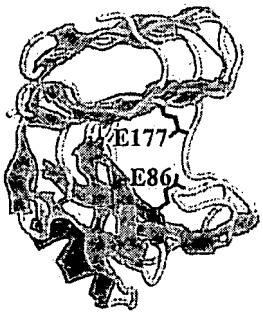
Each Family 11 xylanase three-dimensional structure is identified with the PDB number (The Protein Data Bank: http://www.rcsb.org/pdb/home/home.do). Catalytic residues are shown in dark gray stick and identified using the numbering from Figure 1 in the present specification.



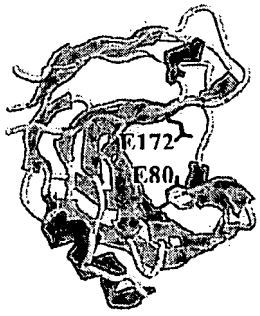
Trichoderma reesei XYN II (PDB 1ENX)



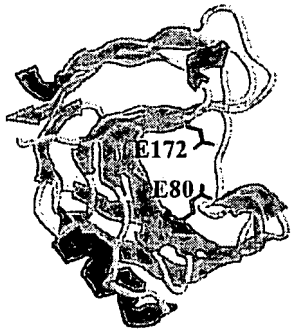
Trichoderma reesei XYN I (PDB 1XYN)



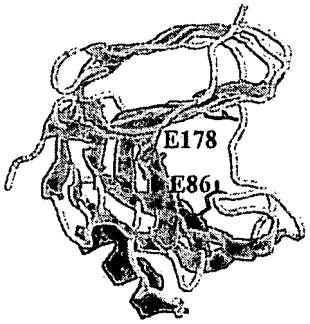
Trichoderma harzianum (PDB 1XND)



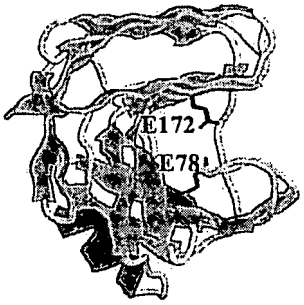
Aspergillus niger (PDB 1UKR)



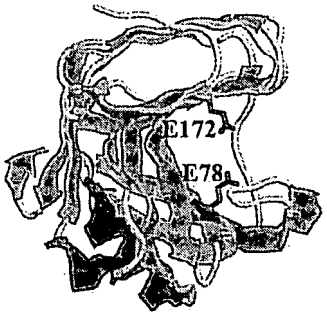
Aspergillus kawachii XynC (PDB 1UKR)



Thermomyces lanuginosus Xyn (PDB YNA)



Bacillus circulans (PDB 1XNB)



Bacillus subtilis (PDB 1IGO)

Structures and mechanisms of glycosyl hydrolases

The wealth of information provided by the recent structure determinations of many different glycosyl hydrolases shows that the substrate specificity and the mode of action of these enzymes are governed by exquisite details of their three-dimensional structures rather than by their global fold.

Structure 15 September 1995, 3:853–859

Carbohydrates show wide stereochemical variation and can be assembled in so many different fashions that there are over 10^{12} possible isomers for a reducing hexasaccharide [1]. Living organisms take advantage of this diversity by using oligosaccharides and polysaccharides for a multitude of biological functions, from storage and structure to highly specific signalling roles. Selective hydrolysis of glycosidic bonds is therefore crucial for energy uptake, cell wall expansion and degradation, and turnover of signalling molecules. As a consequence of saccharide diversity, there is great variety amongst the enzymes that hydrolyze glycosidic bonds, the O-glycosyl hydrolases (EC 3.2.1.x). Heritable deficiencies in glycosyl hydrolases, for example lactose intolerance [2] and a number of lysosomal storage diseases [3], are among the most frequent genetically based syndromes in man.

Mechanisms

Enzymatic hydrolysis of the glycosidic bond takes place via general acid catalysis that requires two critical residues: a proton donor and a nucleophile/base [4,5] (Fig. 1). This hydrolysis occurs via two major mechanisms giving rise to either an overall retention, or an inversion, of anomeric configuration [4]. In both the retaining and the inverting mechanisms, the position of the proton donor is identical, in other words it is within hydrogen-bonding distance of the glycosidic oxygen. In retaining enzymes, the nucleophilic catalytic base is in close vicinity of the sugar anomeric carbon. This base, however, is more distant in inverting enzymes which must accommodate a water molecule between the base and the sugar. This difference results in an average distance between the two catalytic residues of ~5.5 Å in retaining enzymes as opposed to ~10 Å in inverting enzymes [6].

Lysozymes were the first glycosyl hydrolases to have their three-dimensional (3D) structures solved [7,8]. The two catalytic amino acids were identified as aspartate and glutamate residues. In most glycosyl hydrolases studied since, only aspartate and/or glutamate residues have been found to perform catalysis. Recent data, however, suggest that other residues may sometimes be involved in glycosidic bond cleavage. Typical examples are viral neuraminidase and bacterial sialidase, where the transition state is thought to be stabilized with the help of a tyrosine [9,10].

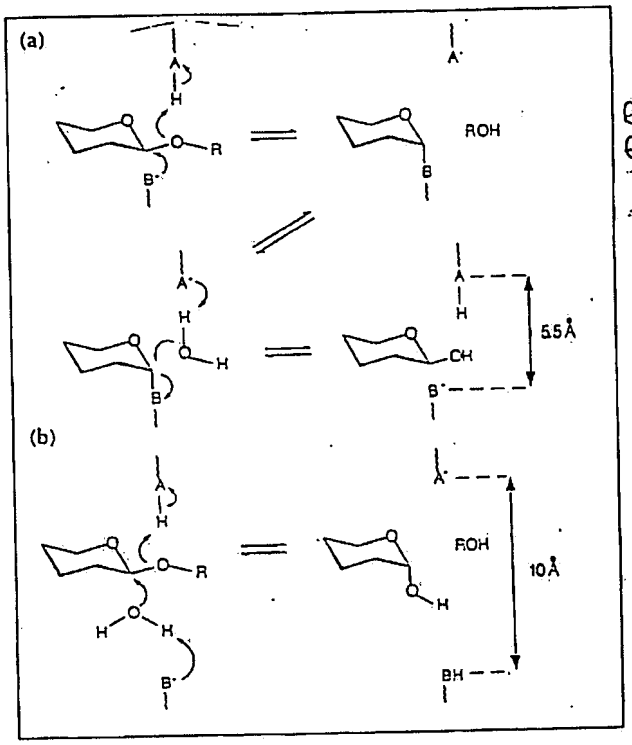


Fig. 1. The two major mechanisms of enzymatic glycosidic bond hydrolysis as first proposed by Koshland [4]. (a) The retaining mechanism, in which the glycosidic oxygen is protonated by the acid catalyst (AH) and nucleophilic assistance to aglycon departure is provided by the base B^-. The resulting glycosyl enzyme is hydrolyzed by a water molecule and this second nucleophilic substitution at the anomeric carbon generates a product with the same stereochemistry as the substrate. (b) The inverting mechanism, in which protonation of the glycosidic oxygen and aglycon departure are accompanied by a concomitant attack of a water molecule that is activated by the base residue (B^-). This single nucleophilic substitution yields a product with opposite stereochemistry to the substrate.

Enzymatic hydrolysis of certain oligosaccharides may also take advantage of the natural chemistry of the substrate. Participation of the substrate C2 acetamido group in catalysis by lysozymes and some retaining chitinases has been suggested from several underrated sources. Hydrolysis by hen egg white lysozyme (HEWL) of *N*-acetylchitobioside substrates with either a C2 hydroxyl or C2 acetamido group has been studied in detail [11].

Removal of the C2 acetamido group leads to a reduction in k_{cat} of at least 100-fold, a factor which is similar, if not greater, than the effect of mutation of the proposed catalytic base Asp52. Perhaps the most serious evidence for the role of the acetamido substituent at C2, both in the enhancement of the catalytic rate and the retention of the anomeric configuration during hydrolysis, comes from solution studies on *N*-acetylglucosamine-containing saccharides [12,13]. Hydrolysis of these sugars, in solution, not only goes faster than that of C2 hydroxyl-substituted sugars, but with overall retention of anomeric configuration, suggesting that, 'on enzyme', a catalytic base, such as Asp52 in HEWL, is not necessarily required. These observations may also help to explain the apparent anomaly that some enzymes known to catalyze the cleavage of β1,4 bonds adjacent to C2 acetamido-substituted saccharides, whose structures have recently been determined, appear to lack a suitable catalytic base located in an appropriate position. Soluble lytic transglycosylase [14], goose lysozyme [15], hevamine [16], and others fall into this category.

Glycosyl hydrolases have developed various ways to lower the energy barrier of the hydrolysis reaction, such as substrate distortion into a sofa or half-chair conformation [17,18]. It is believed that protonation of the glycosidic bond is accompanied by a substantial lengthening of this bond [19]. Oligosaccharide complexes of the endoglucanase V from *Humicola insolens* suggest that this enzyme has evolved a 'stretched' subsite at the catalysis point in order to favour binding of this elongated transition state over that of the ground state substrate molecule (G Davies, S Tolley, B Henrissat, C Hjort and M Schülein, unpublished data).

Families and folds

A classification of glycosyl hydrolases into more than 45 families, on the basis of similarities in amino acid sequence, has recently been proposed [20,21], with the prospect that this may facilitate the derivation of useful information on the structure and function of these enzymes. Underlying this classification was the idea that proteins in a given family would have a fold sufficiently similar to allow homology modelling. In this classification, enzymes with different substrate specificities are sometimes found in the same family, indicating an evolutionary divergence to acquire new specificities, as is found in, for example, families 1, 13 and 16. On the other hand, enzymes that hydrolyze the same substrate are sometimes found in different families. For example, cellulases are found in 11 families. In other words, the sequence, and hence structural, classification differs significantly from that of the International Union of Biochemistry (IUB) nomenclature of enzymes, which is based mostly on substrate specificity.

Because the 3D structures of proteins are more highly conserved than their sequences, several sequence-based families may have related folds. For instance, a structural similarity was suggested for family 11 xylanases and

Table 1. Structures and mechanisms in various families of glycosyl hydrolases.

Family	Enzyme	Organism	EC number	PDB* code	Mechanism	Reference
1	β-glucosidase	*Trifolium repens*	3.2.1.21	–	retaining	[31]
2	β-galactosidase	*Escherichia coli*	3.2.1.23	1BGL	retaining	[32]
5	endoglucanase A	*Clostridium cellulolyticum*	3.2.1.4	–	retaining	[33]
6	cellobiohydrolase II	*Trichoderma reesei*	3.2.1.91	3CBH	inverting	[28]
	endoglucanase	*Thermononospora fusca*	3.2.1.4	1TML	inverting	[34]
7	cellobiohydrolase I	*Trichoderma reesei*	3.2.1.91	1CEL	retaining	[35]
	endoglucanase I	*Humicola insolens*	3.2.1.4	–	retaining	[Davies, unpublished]
9	endoglucanase D	*Clostridium thermocellum*	3.2.1.4	–	inverting	[36]
10	xylanase A	*Streptomyces lividans*	3.2.1.8	1XAS	retaining	[37]
11	xylanase	*Bacillus circulans*	3.2.1.8	1BCX	retaining	[38]
13	α-amylase	*Aspergillus oryzae*	3.2.1.1	6TAA	retaining	[39]
14	β-amylase	*Glycine max*	3.2.1.2	1BTC	inverting	[40]
15	glucoamylase	*Aspergillus awamori*	3.2.1.3	3GLY	inverting	[41]
16	β-1,3-1,4-glucanase	*Bacillus sp.*	3.2.1.73	1BYH	retaining	[42]
17	β-1,3-1,4-glucanase	*Hordeum vulgare*	3.2.1.73	1GHR	unknown†	[43]
18	chitinase	*Serratia marcescens*	3.2.1.14	1CTN	retaining	[44]
19	chitinase	*Hordeum vulgare*	3.2.1.14	1BAA	inverting	[45]
20	chitobiase	*Serratia marcescens*	3.2.1.52	–	retaining	‡
22	lysozyme	Hen egg white	3.2.1.17	1HEL	retaining	[7]
23	lysozyme	Goose	3.2.1.17	153L	unknown	[15]
24	lysozyme	Bacteriophage T4	3.2.1.17	1LYD	unknown	[46]
33	sialidase	*Salmonella typhimurium*	3.2.1.18	2SIL	retaining	[10]
34	neuraminidase	Influenza virus B	3.2.1.18	1NSB	retaining	[47]
45	endoglucanase V	*Humicola insolens*	3.2.1.4	1ENG	inverting	[48]

*Protein Data Bank. †This family is predicted to have a retaining mechanism [24,25]. ‡I Tews, Z Dauter, KS Wilson & CE Vorgias, [abstract 038], 4th European Workshop on Crystallography of Biological Macromolecules, Como, Italy, May 1995.

family 12 cellulases [22], whereas family 7 cellulases have been found to have an arrangement of catalytic residues and a fold similar to those of the β-1,3-glucanases and β-1,3-1,4-glucanases of family 16 [23]. More recently, families 1, 2, 5, 10, 17, 30, 35, 39 and 42 were proposed to have evolved from a common ancestor [24,25]. All of these family groupings substantiate the strict conservation of the catalytic machinery and mechanism during evolution. A grouping indicating structural similarity for various lysozymes (families 22, 23 and 24) and family 19 plant chitinases has also been proposed [26]. This grouping, however, includes enzymes known to operate with retention of configuration, such as HEWL, and some with a substantially different arrangement of catalytic amino acids, such as family 19 plant chitinases, which are inverting enzymes [27].

Table 1 reports the various glycosyl hydrolases families for which at least one 3D structure has been determined, together with the mechanism of glycosidic bond hydrolysis, in cases where it is known. The substrates for these enzymes are shown in Figure 2. Many glycosyl hydrolases have a modular structure consisting of a catalytic domain and one or more non-catalytic domains, some of which are involved in substrate binding, but most of which have unknown functions. Figure 3 shows the main folds found in the catalytic domains of selected glycosyl hydrolases and, for families 6 and 7, the comparison of the structures of cellobiohydrolases with those of the corresponding endoglucanases (see below).

Active-site topologies

Although many protein folds are represented in the 22 families for which a 3D structure is known, the overall topologies of the active sites fall into only three general classes, regardless of whether the enzyme is inverting or retaining. These three topologies (Fig. 4) can, in principle, be built on the same fold, with the same catalytic residues.

Pocket or crater

This topology (Fig. 4a) is optimal for the recognition of a saccharide non-reducing extremity and is encountered in monosaccharidases such as β-galactosidase, β-glucosidase, sialidase and neuraminidase, and in exopolysaccharidases such as glucoamylase and β-amylase. Such exopolysaccharidases are adapted to substrates having a large number of available chain ends, such as native starch granules, whose radial structure exposes all the non-reducing chain ends at the surface. On the other hand, these enzymes are not very efficient for fibrous substrates such as native cellulose, which has almost no free chain ends.

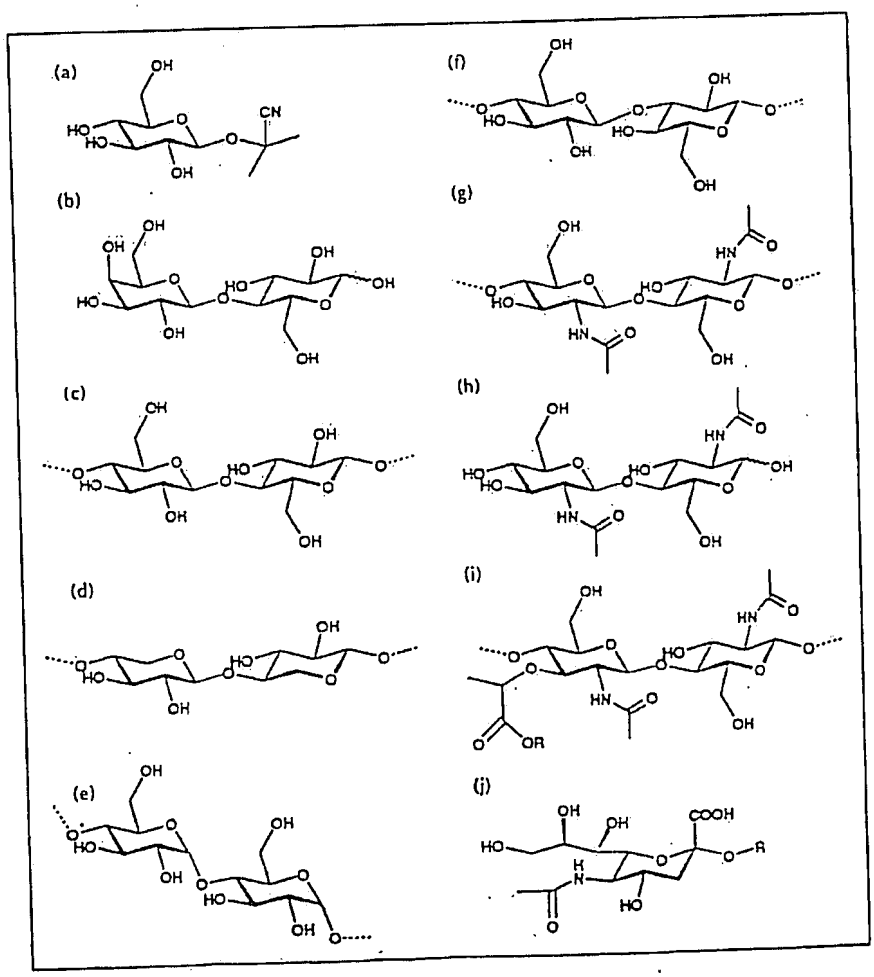


Fig. 2. Structures of the substrates for the enzymes given in Table 1. The hydrolyzed bond is shown in red. (a) Linamarin (substrate for cyanogenic β-glucosidase; family 1). (b) Lactose (β-galactosidase, family 2). (c) Cellulose (endoglucanases and cellobiohydrolases, families 5, 6, 7, 9 and 45). (d) Xylan (xylanases, families 10 and 11). (e) Amylose (α-amylase, family 13; β-amylase, family 14; and glucoamylase, family 15). (f) Mixed β-1,3-1,4-glucan with the bond hydrolyzed by β-1,3-glucanases (family 16) in red and that cleaved by β-1,3-1,4-glucanases (family 17) in cyan. (g) Chitin (chitinases in families 18 and 19, family 22 lysozymes). (h) Chitobiose (chitobiase, family 20). (i) Bacterial cell wall polymer consisting of alternating *N*-acetylmuramic acid and *N*-acetylglucosamine units (lysozymes in families 22, 23 and 24). (j) Glycoconjugate-linked sialic acid (sialidase, family 33 and neuraminidase, family 34).

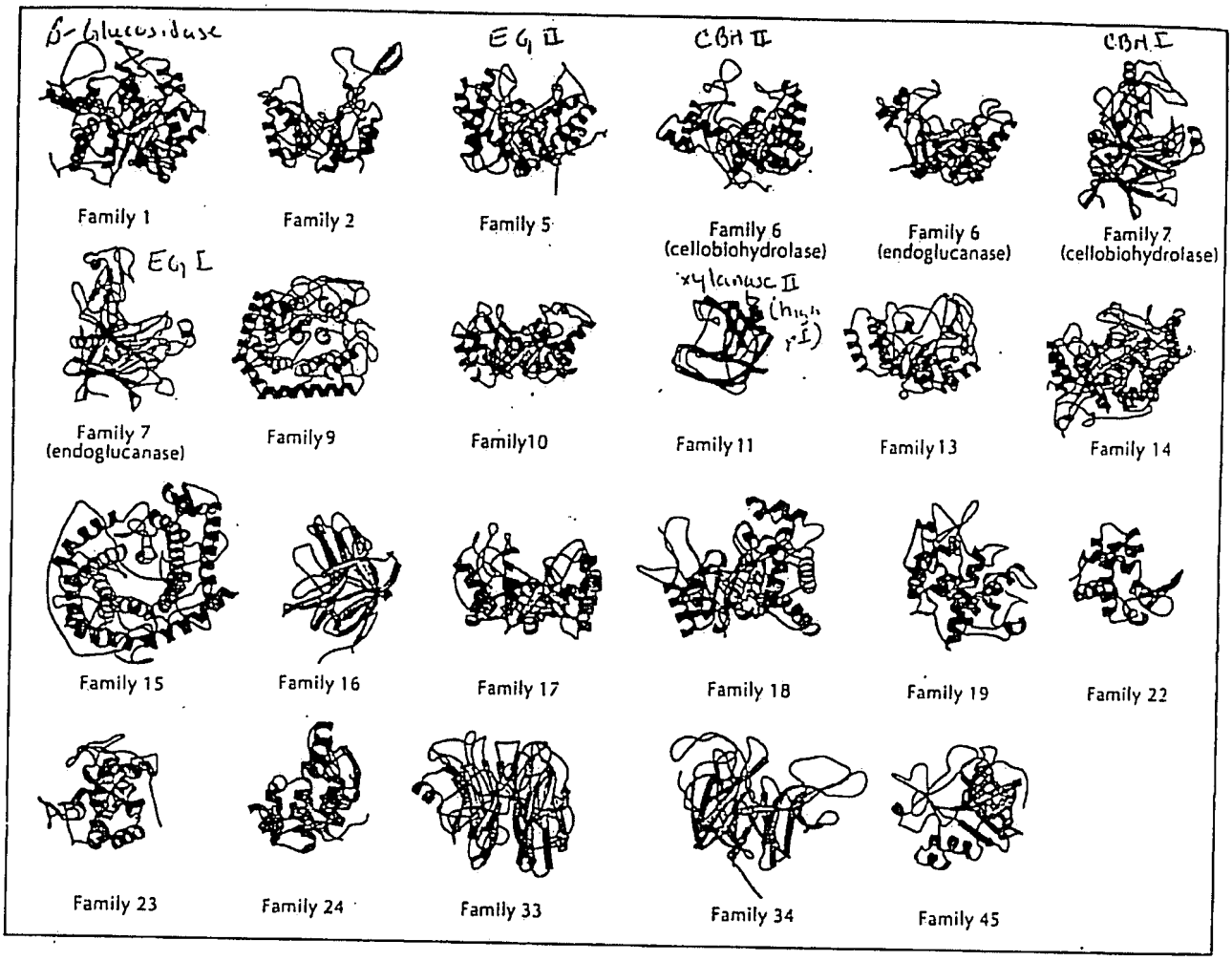


Fig. 3. Ribbon representation of the main fold of the catalytic domain in various glycosyl hydrolase families (see Table 1 and text). β strands are shown in cyan and α helices in red. (Figure produced using the program MOLSCRIPT [49].)

Cleft or groove

This 'open' structure (Fig. 4b) allows a random binding of several sugar units in polymeric substrates and is commonly found in endo-acting polysaccharidases such as lysozymes, endocellulases, chitinases, α-amylases, xylanases, β-1,3-1,4-glucanases and β-1,3-glucanases.

Tunnel

This topology (Fig. 4c) arises from the previous one when the protein evolves long loops that cover part of the cleft. Found so far only in cellobiohydrolases, the resulting tunnel enables a polysaccharide chain to be threaded through it [28]. A comparison of the cellobiohydrolases in families 6 and 7 with the corresponding endoglucanases is shown in Figure 3. The loops that cause the catalytic centres of cellobiohydrolases to lie within enclosed tunnels can be seen clearly and, for the family 6 enzymes, are also illustrated by the surface representations in Figures 4b and 4c. This topology allows these enzymes to release the product while remaining firmly bound to the polysaccharide chain, thereby creating the conditions for processivity (Fig. 5). It remains unclear at present whether the substrate initially penetrates the active site in an 'exo' fashion by one of the two entrances of the tunnel or whether the loops that close the active site can 'open' occasionally to allow a random binding followed by the processive action. In either case it should be noted that, depending on the mechanism (inverting or retaining) and the exact position of the cleavage point with respect to the several subsites, the directionality of the enzyme motion along the chain may change. For instance, cellobiohydrolase II of *Trichoderma reesei* proceeds towards the reducing end of cellulose, whereas the reverse was suggested for cellobiohydrolase I from the same organism [23]. Processivity is probably a key factor for the efficient enzymatic degradation of insoluble microcrystalline cellulose.

Concluding remarks

Orengo *et al.* [29] have shown that certain protein folds (superfolds) occur more often than others. More precisely, only nine superfolds are sufficient to describe the folding in ~30% of all proteins and it is thought that the total number of protein folds is not more than a few thousand [29,30]. So far, from the 22 families of glycosyl hydrolases for which a 3D structure has been

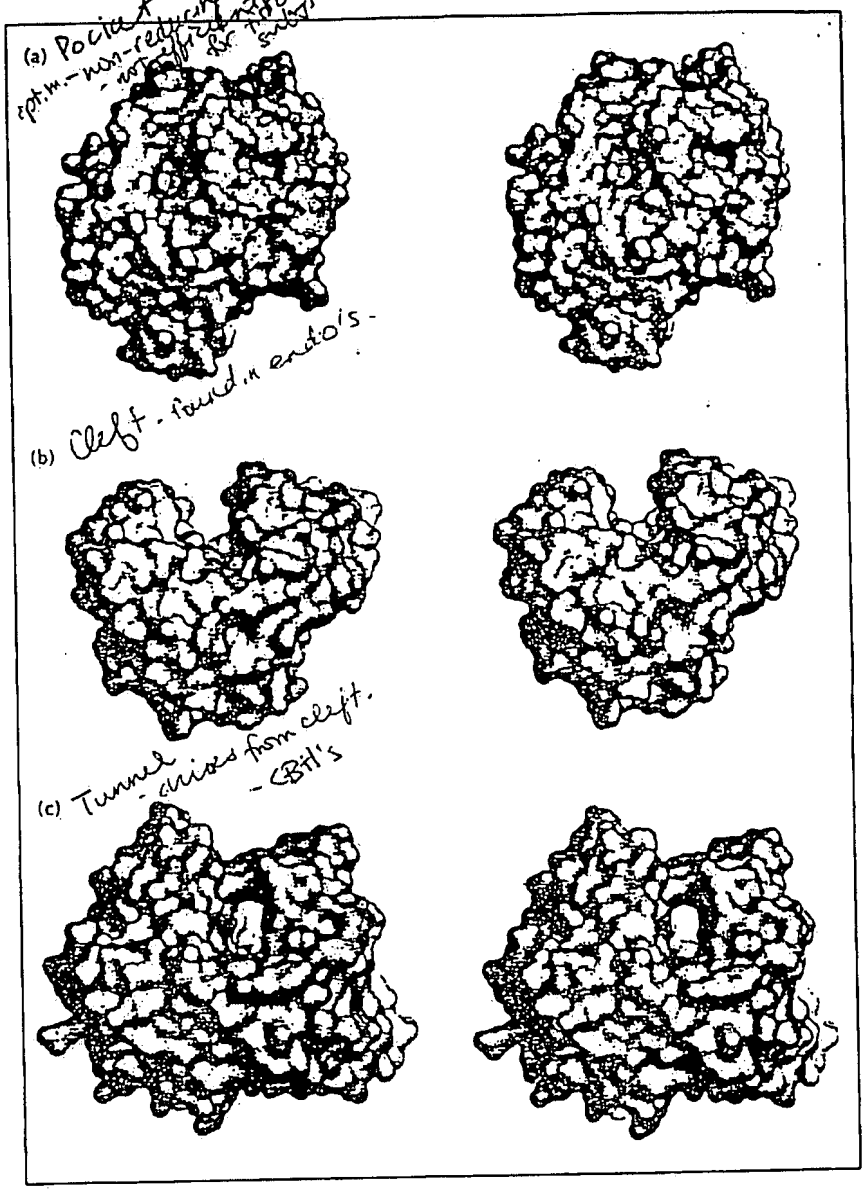


Fig. 4. The three types of active site found in glycosyl hydrolases. (a) The pocket (glucoamylase from *A. awamori*). (b) The cleft (endoglucanase E2 from *T. fusca*). (c) The tunnel (cellobiohydrolase II from *T. reesei*). The proposed catalytic residues are shaded in red. (Molecular surface diagrams were prepared using the MOLVIEWER program [M Hartshorn, unpublished program].)

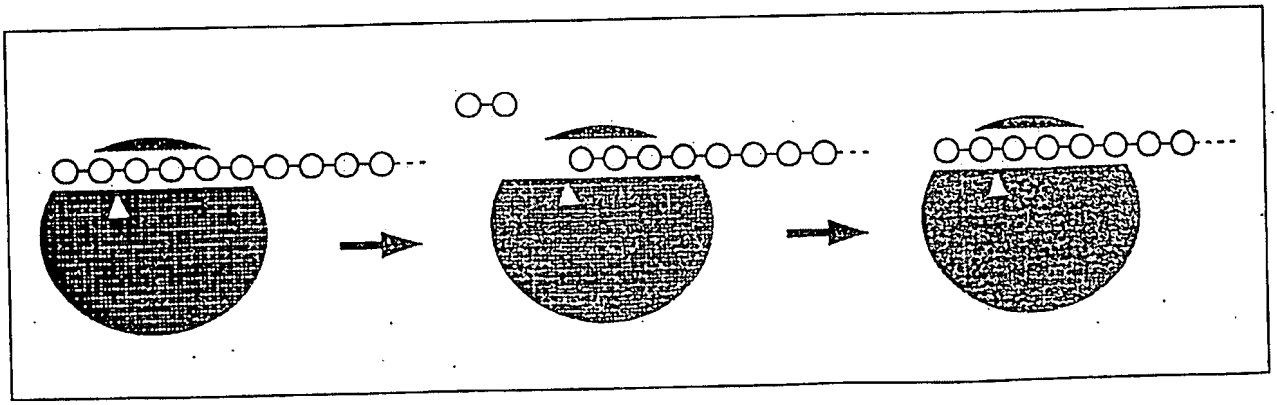


Fig. 5. The mechanism of processivity of cellobiohydrolases. Once a disaccharide product is liberated (shown as two linked circles), the enzyme remains bound to the polysaccharide chain by several subsites, the 'lid' closing the active site and, for retaining enzymes, the glycosyl enzyme. The two empty sites, and perhaps other factors such as loop movements, provide the driving force for enzyme motion along the chain (or chain threading along the enzyme's active site) by two sugar units. Hydrolysis can then proceed iteratively until enzyme movement is stopped by steric factors, or until the loops that close the active site move and release the polysaccharide chain.

determined, nine are of the TIM barrel type (Table 1). In addition, detailed sequence comparisons have suggested that several other families also share the same fold, implying that, for glycosyl hydrolases, the bias towards this fold could be even more important [24].

There are presently 52 families in the classification based on sequence similarities (B Henrissat and A Bairoch, unpublished data); 3D structures are known for about 40% of these families. The first enzyme structure to be solved, more than 30 years ago, was that of hen egg white lysozyme, a glycosyl hydrolase. At the present pace of structural investigations, the 3D fold for all of the remaining families could be determined within a few years.

Acknowledgements: We thank those who gave coordinates not yet available from the Protein Data Bank: Tracey Barrett (Mill Hill) for the cyanogenic β-glucosidase of *Trifolium repens*; Zygmunt and Ulla Derewenda (University of Alberta) for xylanase A from *Streptomyces lividans*; Valérie Ducros (CNRS Marseille) for endoglucanase A from *Clostridium cellulolyticum*; Michel Juy (CNRS Marseille) for endoglucanase D from *Clostridium thermocellum*; and Brian Matthews (University of Oregon) for β-galactosidase from *E. coli* and for goose lysozyme. We would like to thank Mike Hartshorn (University of York) for help with the MOLVIEWER program and the production of figures. This work has been funded by the BIO2 CT 94 3018 research grant from the European Community.

References

1. Laine, R.A. (1994). A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the *Isomer Barrier* to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* 4, 759–767.
2. Semenza, G. & Auricchio, S. (1989). Small-intestinal disaccharidases. In *The Metabolic Basis of Inherited Diseases*. (Scrives, C.R., Beaudet, A.L. & Valle, V.S., eds), pp. 2975–2997, McGraw-Hill, NY.
3. Neufeld, E.F. (1991). Lysosomal storage diseases. *Annu. Rev. Biochem.* 60, 257–280.
4. Koshland, D.E. (1953). Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. Camb. Philos. Soc.* 28, 416–436.
5. Sinnott, M.L. (1990). Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev.* 90, 1171–1202.
6. McCarter, J.D. & Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* 4, 885–892.
7. Blake, C.C.F., Koenig, D.F., Mair, G.A., North, A.C.T., Phillips, D.C. & Sarma, V.R. (1965). Structure of hen egg white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* 206, 757–763.
8. Matthews, B.W. & Remington, S.J. (1974). The three-dimensional structure of lysozyme from bacteriophage T4. *Proc. Natl. Acad. Sci. USA* 71, 4178–4182.
9. Burmeister, W., Henrissat, B., Bosso, C., Cusack, S. & Ruigrok, R. (1993). Influenza B neuraminidase can synthesize its own inhibitor. *Structure* 1, 19–26.
10. Crennel, S.J., Garman, E.F., Laver, W.G., Vimr, E.R. & Taylor, G.L. (1993). Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase. *Proc. Natl. Acad. Sci. USA* 90, 9852–9856.
11. Lowe, G. & Sheppard, G. (1968). Acetamido-group participation in lysozyme catalysis. *J. Chem. Soc., Chem. Commun.* 529–530.
12. Capon, B. (1969). Mechanisms in carbohydrate chemistry. *Chem. Rev.* 69, 407–498.
13. Piszkiewicz, D. & Bruice, T.C. (1968). Glycoside hydrolysis. II. Intramolecular carboxyl and acetamido group catalysis in β-glycoside hydrolysis. *J. Am. Chem. Soc.* 90, 2156–2163.
14. Thunnissen, A.-M., *et al.*, & Dijkstra, B.W. (1994). Doughnut-shaped structure of a bacterial muramidase revealed by X-ray crystallography. *Nature* 367, 750–753.
15. Weaver, L.H., Grütter, M.G. & Matthews, B.W. (1995). The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the "goose-type" lysozymes lack a catalytic aspartate residue. *J. Mol. Biol.* 245, 54–68.
16. Terwisscha van Scheltinga, A.C., Kalk, K.H., Beintema, J.J. & Dijkstra, B.W. (1994). Crystal structure of hevamine, a plant defence protein with chitinase and lysozyme activity, and its complex with an inhibitor. *Structure* 2, 1181–1189.
17. Strynadka, N.C.J. & James, M.N.G. (1991). Lysozyme revisited: crystallographic evidence for distortion of an *N*-acetylmuramic acid residue bound in site D. *J. Mol. Biol.* 220, 401–424.
18. Kuroki, R., Weaver, L.H. & Matthews, B.W. (1993). A covalent enzyme–substrate intermediate with saccharide distortion in a mutant T4 lysozyme. *Science* 262, 2030–2033.
19. Tanaka, Y., Tao, W., Blanchard, J.S. & Hehre, E.J. (1994). Transition state structures for the hydrolysis of α-D-glucopyranosyl fluoride by retaining and inverting reactions of glycosylases. *J. Biol. Chem.* 269, 32306–32312.
20. Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarity. *Biochem. J.* 280, 309–316.
21. Henrissat, B. & Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 293, 781–788.
22. Törrönen, A., Kubicek, C.P. & Henrissat, B. (1993). Amino acid sequence similarities between low molecular weight endo-1,4-β-xylanases and family H cellulases revealed by clustering analysis. *FEBS Lett.* 321, 135–139.
23. Divne, C., *et al.*, & Jones, T.A. (1994). The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* 265, 524–528.
24. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P. & Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. USA* 92, 7090–7094.
25. Jenkins, J., Lo Leggio, L., Harris, G. & Pickersgill, R. (1995). β-glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily with 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. *FEBS Lett.* 362, 281–285.
26. Holm, L. & Sander, C. (1994). Structural similarity of plant chitinase and lysozymes from animals and phage. *FEBS Lett.* 340, 129–132.
27. Fukamizo, T., Koga, D. & Goto, S. (1995). Comparative biochemistry of chitinases — anomeric form of the reaction products. *Biosci. Biotechnol. Biochem.* 59, 311–313.
28. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C. & Jones, T.A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* 249, 380–386.
29. Orengo, C., Jones, D.T. & Thornton, J.M. (1994). Protein superfamilies and domain superfolds. *Nature* 372, 631–634.
30. Chothia, C. (1992). One thousand families for the molecular biologist. *Nature* 357, 543–544.
31. Barrett, T., Suresh, C.G., Tolley, S.P., Dodson, E.J. & Hughes, M.A. (1995). The crystal structure of a cyanogenic β-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure* 3, 951–960.
32. Jacobson, R.H., Zhang, X.-J., DuBose, R.F. & Matthews, B.W. (1994). Three-dimensional structure of β-galactosidase from *E. coli*. *Nature* 369, 761–766.
33. Ducros, V., *et al.*, & Haser, R. (1995). Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure* 3, 939–949.
34. Spezio, M., Wilson, D.B. & Karplus, P.A. (1993). Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry* 32, 9906–9916.
35. Divne, C., *et al.*, & Jones, T.A. (1994). The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* 265, 524–528.
36. Juy, M., *et al.*, & Aubert, J.-P. (1992). Three-dimensional structure of a thermostable bacterial cellulase. *Nature* 357, 89–91.
37. Derewenda, U., *et al.*, & Derewenda, Z.S. (1994). Crystal structure, at 2.6 Å resolution, of the *Streptomyces lividans* xylanase A, a member of the F family of β-1,4-D-glycanases. *J. Biol. Chem.* 269, 20811–20814.
38. Wakarchuk, W.W., Campbell, R.L., Sung, W., Davoodi, J. & Yaguchi, M. (1994). Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase. *Protein Sci.* 3, 467–475.
39. Boel, E., *et al.*, & Woldike, H.F. (1990). Calcium binding in α-amylases: an x-ray diffraction study at 2.1-Å resolution of two enzymes from *Aspergillus*. *Biochemistry* 29, 6244–6249.
40. Mikami, B., *et al.*, & Sacchettini, J.C. (1993). The 2.0 Å resolution structure of soybean β-amylase complexed with α-cyclodextrin. *Biochemistry* 32, 6836–6845.
41. Aleshin, A., Golubev, A., Firsov, L.M. & Honzatko, R.B. (1992). Crystal structure of glucoamylase from *Aspergillus awamori* var. *X100* to 2.2 Å resolution. *J. Biol. Chem.* 267, 19291–19298.

42. Keitel, T., Simon, O., Borriss, R. & Heinemann, U. (1993). Molecular and active-site structure of a *Bacillus* 1,3-1,4-β-glucanase. *Proc. Natl. Acad. Sci. USA* 90, 5287–5291.
43. Varghese, J.N., Garrett, T.P.J., Colman, P.M., Chen, L., Høj, P.B. & Fincher, G.B. (1994). Three-dimensional structure of two plant β-glucan endohydrolases with distinct substrate specificities. *Proc. Natl. Acad. Sci. USA* 91, 2785–2789.
44. Perrakis, A., *et al.*, & Vorgias, C.E. (1994). Crystal structure of a bacterial chitinase at 2.3 Å resolution. *Structure* 2, 1169–1180.
45. Hart, P.J., Monzingo, A.F., Ready, M.P., Ernst, S.R. & Robertus, J.D. (1993). Crystal structure of an endochitinase from *Hordeum vulgare* L. seeds. *J. Mol. Biol.* 229, 189–193.
46. Weaver, L.H. & Matthews, B.W. (1987). Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. Mol. Biol.* 193, 189–200.
47. Varghese, J.N., Laver, W.G. & Colman, P.M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature* 303, 35–40.
48. Davies, G.J., *et al.*, & Schülein, M. (1993). Structure and function of endoglucanase V. *Nature* 365, 362–364.
49. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* 24, 946–950.

Gideon Davies, Department of Chemistry, University of York, Heslington, York, YO1 5DD, UK and Bernard Henrissat, Centre de Recherches sur les Macromolécules Végétales, CNRS, BP 53, F-38041 Grenoble Cedex 9, France.

ELSEVIER

Journal of Biotechnology 95 (2002) 109–131

JOURNAL OF Biotechnology

www.elsevier.com/locate/jbiotec

The endoxylanases from family 11: computer analysis of protein sequences reveals important structural and phylogenetic relationships

Amalia Sapag [a,1], Johan Wouters [b], Christophe Lambert [b], Pablo de Ioannes [a], Jaime Eyzaguirre [a,c,*], Eric Depiereux [b]

[a] *Laboratorio de Bioquímica, Departamento de Genética Molecular y Microbiología, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile*

[b] *Department of Biology, Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium*

[c] *Escuela de Artes Liberales, Universidad Nacional Andrés Bello, Santiago, Chile*

Received 5 June 2001; received in revised form 5 November 2001; accepted 20 December 2001

Abstract

Eighty-two amino acid sequences of the catalytic domains of mature endoxylanases belonging to family 11 have been aligned using the programs MATCHBOX and CLUSTAL. The sequences range in length from 175 to 233 residues. The two glutamates acting as catalytic residues are conserved in all sequences. A very good correlation is found between the presence (at position 100) of an asparagine in the so-called 'alkaline' xylanases, or an aspartic acid in those with a more acidic pH optimum. Four boxes defining segments of highest similarity were detected; they correspond to regions of defined secondary structure: B5, B6, B8 and the carboxyl end of the alpha helix, respectively. Cysteine residues are not common in these sequences (0.7% of all residues), and disulfide bridges are not important in explaining the stability of several thermophilic xylanases. The alignment allows the classification of the enzymes in groups according to sequence similarity. Fungal and bacterial enzymes were found to form mostly separate clusters of higher similarity.

Keywords: Family 11 endoxylanases; Factor analysis classification; Sequence alignment; Structural relationships

1. Introduction

Xylan, a heteroglycan, is the main constituent of plant hemicelluloses. It is composed of a linear chain of xylose residues linked by $\beta(1 \rightarrow 4)$ glycosidic bonds, with a variety of substituents, depending on its source (Joseleau et al., 1992). Biodegradation of xylan is accomplished by a complex set of enzymes generically called 'xy-

* Corresponding author. Tel.: +562-686-2664; fax: +562-222-5515.

E-mail address: jeyzag@genes.bio.puc.cl (J. Eyzaguirre).

[1] Present address: Laboratorio de Farmacoterapia Génica, Facultad de Ciencias Químicas y Farmacéuticas, Unversidad de Chile, Santiago, Chile.

0168-1656/02/$ - see front matter
PII: S0168-1656(02)00002-0

lanases', which are produced by fungi and bacteria (Biely, 1985).

The hydrolysis of the xylan backbone is performed mainly by the action of the endoxylanases (E.C. 3.2.1.8), which liberate xylooligosaccharides of different length (Biely, 1985). A large number of endoxylanases have been described, purified, characterized and sequenced from different microorganisms (Sunna and Antranikian, 1997), and based on sequence similarities and hydrophobic cluster analysis, they have been grouped in families 10 (or F) and 11 (or G) of the glycosyl hydrolases (Gilkes et al., 1991; Henrissat and Davies, 1997). Family 10 xylanases have a catalytic domain with molecular weights in the range of 32–39 K. The structure of the catalytic domain of several family 10 xylanases has been determined by X-ray crystallography and it consists of an eightfold β/α barrel (Harris et al., 1996). These enzymes show a greater catalytic versatility than family 11 endoxylanases (Biely et al., 1997). Family 11 xylanases, on the other hand, have a much smaller catalytic domain (around 20 kDa), with an all β-strand sandwich fold structure (Himmel et al., 1997). Some enzymes of family 10 and family 11 possess, in addition, cellulose or xylan binding domains (Törrönen and Rouvinen, 1997).

The endoxylanases have been found to be useful for several different biotechnological applications (Section 3.4). Therefore, for the design and protein engineering of endoxylanases, a good knowledge of their sequence and structure is of great importance.

For this purpose, a comparative analysis of the sequences of 82 catalytic domains of family 11 glycanases has been performed in this work. The large number of sequences available allows a fine analysis of sequence similarities, with the purpose of establishing structural relationships, determine location of conserved (and possibly essential) sequences and establish phylogenetic relationships.

2. Materials and methods

Sequences were obtained from the literature and from the public databases (SwissProt, GenBank and CAZy (http://afmb.cnrs-mrs.fr/~pedro/CAZY/ghf.html)). The 82 sequences were aligned by the method described below.

To define conserved residues in a reliable alignment of numerous sequences of very different length and of poor similarity is not trivial (de Fays et al., 1999). Although alignments are widely used, the rate of false positive functional similarities deduced from skewed aligned positions is not easily overcome (Briffeuil et al., 1998). In this study, an original methodology has been applied to obtain reliable alignments allowing a precise delineation of the corresponding relevant residues in the different sequences, despite the fact that some sequences share a very low percentage of identity (10%).

The first step of this methodology is based on the local alignment method MATCHBOX (Depiereux and Feytmans, 1992). Default parameters were used to delineate conserved boxes (defining Predicted Structural Conserved Regions, PSCR) with an estimated level of confidence (from 1 to 9) at each position of the alignment (Depiereux et al., 1997). In this work, PSCR's are defined for reliability scores below 5 on more than 2 non-redundant sequences, which corresponds to a rate of confidence of over 90%. This ensures less than 10% of 'false positive' aligned positions, a remarkable rate at this level of low similarity.

The second step of the procedure is based on the global alignment method CLUSTAL (Higgins et al., 1992). It is used to align the regions inserted between the anchor points defined above. Thus, the final alignment obtained by this original procedure takes advantage of both approaches: MATCHBOX for definition of the boxes (PSCR's), and CLUSTAL for the multiple alignment of the segments outside the boxes.

The proposed classification of the family 11 endoxylanases is based on a principal coordinates analysis computed from a similarity matrix between the sequences (Depiereux and Feytmans, 1991). The method takes advantage of a grouping of the sequences independent of any alignment steps and thus avoids a misclassification due to misalignments. Briefly, the similarity coefficient is based on matches, for a given threshold, between short unaligned segments of the sequences compared; eigenvectors and eigenvalues are computed

by matrix diagonalization, and sequences are plotted for factors 2 and 3, factor 1 being poorly discriminant. Eigenvector coordinates have been submitted to a cluster analysis, Ward's method, run on STATISTICA 4.1 (Statsoft, Inc) and plotted in an EXCEL spreadsheet.

PSCR delineation and factor analysis have been performed by the programs Align and Explore on the MATCHBOX server http://www.fundp.ac.be/sciences/biologie/bms/matchbox_submit.shtml and global alignment on the Clustal server http://dot.imgen.bcm.tmc.edu:9331/multi-align/Options/clustalw.html.

3. Results and discussion

Table 1 lists the 82 enzymes used in this study. It includes 36 sequences from bacteria, 43 from fungi, 2 from protozoa and 1 from insects. Twenty-two bacterial and 8 fungal enzymes are multidomain proteins; they correspond mainly to anaerobic microorganisms. Only the sequence of the catalytic domain of the mature enzymes is considered in this analysis. The amino terminus of the mature enzymes was taken to be either the experimentally determined terminus, when reported in the literature, or the predicted terminus, determined by means of the program SIGNALP (Nielsen et al., 1997). The enzymes from *Fibrobacter succinogenes*, *Neocallimastix frontalis* and *Neocallimastix patriciarum* have two family 11 catalytic domains each; these domains are considered separately in the analysis. The size of the catalytic domain of the multidomain enzymes is somewhat arbitrary; in Fig. 1 the sequences belonging to these enzymes are marked with a star at the beginning and/or end to indicate this fact.

The enzymes analyzed (not considering the multidomain proteins) range in molecular mass from 19 035 (*Trichoderma reesei*) to 25 908 (*Clostridium acetobutylicum*), and in sequence length from 175 (*Polypastron multivesiculatum* Xyn A) to 233 amino acid residues (*C. acetobutylicum*). A shorter sequence (Xyn 4 from *Aspergillus niger*; 153 residues, accession number U39785) was not included in this study because it aligned poorly to the other sequences due to its shorter length, although it is considered part of family 11 by CAZy.

Table 2 shows the pI's, pH and T° optima of those enzymes from the above list which have been purified and characterized. The pI values given are values measured in the laboratory; no theoretical estimations have been included. A wide range of pI values has been found: from a low of 3.5 for *Aspergillus kawachii* and *Aspergillus nidulans* to a high of > 10.25 for *Streptomyces lividans* xylanases. Bacterial enzymes show pH optima ranging from 5.5 to 7, while those from fungi show a much wider range (from pH 2.0 to 8.0), the majority having acidic pH optima.

The three-dimensional structure of family 11 endoxylanases has been determined for several enzymes, from both bacteria and fungi (Table 3). The catalytic domain folds into two β sheets (A and B) constituted mostly by antiparallel β strands and one short alpha helix and resembles a partially closed right hand (Törrönen and Rouvinen, 1997). The loop between strands B7 and B8 forms the 'thumb' and the loop linking strands B6 and B9 is the 'cord'. A great similarity is found in all these structures. Fig. 2 presents a topology diagram showing the sequence of the secondary structure elements found in these enzymes. Two glutamic acid residues have been found to be catalytically essential and are located in strands B4 and B6, respectively (Törrönen and Rouvinen, 1997).

3.1. Multiple sequence alignment

The original alignment methodology used in this work is aimed at delineating structural features shared by all the xylanase sequences included in the analysis. First, boxes obtained from the MATCHBOX alignment program have been shown to accurately outline conserved structural motifs (De Bolle et al., 1995; Vinals et al., 1995; Bertrand et al., 1997, 1998; Depiereux et al., 1997). The fact that these predicted structurally conserved regions actually reflect robust structural similarities among the endoxylanases has been checked by comparison of the three-dimensional structures of the 11 xylanases of known crystallographic structure (Table 3) using the HOMOLOGY

Table 1
Family 11 xylanases used in this study

Organism	Protein	Gene	Accession No. (GenBank)	Catalytic domain		References
				Length[a]	MW	
Bacteria						
Aeromonas caviae	Xylanase I	*xynA*	D32065	183	20 212	Kubata et al. (1997)
Bacillus agaradhaerens	Xylanase		A48223	221	24 681	Sabini et al. (1999)
Bacillus circulans	XLNA	*xlnA*	X07723	185	20 382	Yang et al. (1988)
Bacillus pumilus	XYNA	*xynA*	X00660	201	22 515	Fukusaki et al. (1984)
Bacillus sp.	Xylanase Y[b]	*xynY*	S51779	200	22 201	Yu et al. (1993)
Bacillus sp.	Xylanase S	*xynS*	X59058	185	20 364	Yu et al. (1993)
Bacillus sp.	Xylanase	*xynA*	U51675	185	20 220	Jeong et al. (1998)
Bacillus sp. D3	Xylanase	*xyn*		182	20 683	Harris et al. (1997)
Bacillus sp. 41 M-1	Xylanase J[b]	*xynJ*	AB029319	199	22 098	Nakai et al. (1994)
Bacillus stearothermophilus	XYNA	*xynA*	U15985	191	21 083	Cho and Choi (1995)
Bacillus subtilis	Xylanase A	*xynA*	M36648	185	21 451	Paice et al. (1986)
Caldicellulosiruptor sp. Rt69B.1	Xylanase[b]	*xynD*	AF036925	199	22 158	Morris et al. (1999)
Cellulomonas fimi	XYLD[b]	*xynD*	X76729	198	21 451	Millward-Sadler et al. (1994)
Cellvibrio mixtus	XYLA[b]	*xynA*	Z48925	207	22 550	Millward-Sadler et al. (1995)
Clostridium acetobutylicum	Xylanase B	*xynB*	M31726	233	25 908	Zappe et al. (1990)
Clostridium stercorarium	XynA[b]	*xynA*	D13325	193	22 130	Sakka et al. (1993)
Clostridium thermocellum	Xylanase U[b]	*xynU*	AF047761	204	22 842	Unpublished
Clostridium thermocellum	Xylanase V[b]	*xynV*	AF047761	204	22 787	Unpublished
Clostridium thermocellum	XynA[b]	*xynA*	AB010958	200	22 445	Hayashi et al. (1999)
Clostridium thermocellum	XynB[b]	*xynB*	AB010958	200	22 365	Hayashi et al. (1999)
Dictyoglomus thermophilum	Xylanase B[b]	*xynB*	U76545	198	22 204	Morris et al. (1998)
Fibrobacter succinogenes	XynC (domA)[b]	*xynC*	U01037	234	25 530	Paradis et al. (1993)
Fibrobacter succinogenes	XynC (domB)[b]	*xynC*	U01037	216	24 437	Paradis et al. (1993)
Pseudomonas fluorescens	XYLE[b]	*xynE*	Z48927	202	22 130	Millward-Sadler et al. (1995)
Ruminococcus albus	XynA[b]	*xynA*	U43089	236	26 314	Unpublished
Ruminococcus flavefaciens	XYLA[b]	*xynA*	Z11127	221	24 346	Zhang and Flint (1992)
Ruminococcus flavefaciens	XynB[b]	*xynB*	Z35226	216	24 216	Zhang et al. (1994)
Ruminococcus flavefaciens	XYLD[b]	*xynD*	S61204	213	24 031	Flint et al. (1993)
Ruminococcus sp.	Xylanase 1[b]	*xyn1*	Z49970	213	23 904	Unpublished
Streptomyces lividans	XlnB[b]	*xlnB*	M64552	192	21 064	Shareck et al. (1991)
Streptomyces lividans	XlnC	*xlnC*	M64553	191	20 715	Shareck et al. (1991)
Streptomyces sp. EC3	Xylanase	*xln*	X81045	191	20 931	Mazy-Servais et al. (1996)
Streptomyces sp. S38	Xylanase	*xyl1*	X985518	190	20 585	Georis et al. (1999)
Streptomyces sp. 36a	Xylanase			192	20 973	Nagashima et al. (1989)
Streptomyces thermoviolaceus	STX-II[b]	*stx*-II	D85897	190	20 738	Tsujibo et al. (1997)
Thermomonospora fusca	TfxA[b]	*xynA*	U01242	190	20 900	Irwin et al. (1994)
Fungi						
Ascochyta pisi	Xylanase	*xyl1*	Z68891	208	22 185	Lübeck et al. (1997)
Aspergillus awamori	EXLA	*exlA*	X78115	184	19 876	Hessing et al. (1994)

Table 1 (*Continued*)

Organism	Protein	Gene	Accession No. (GenBank)	Catalytic domain		References
				Length[a]	MW	
Aspergillus kawachii	Xylanase B	*xynB*	P48824[c]	207	22 259	Unpublished
Aspergillus kawachii	XynC	*xynC*	D14848	184	19 876	Ito et al. (1992a)
Aspergillus nidulans	X22	*xlnA*	Z49892	188	20 235	Perez-González et al. (1996)
Aspergillus nidulans	X24	*xlnB*	Z49893	188	20 077	Perez-González et al. (1996)
Aspergillus niger	Xylanase A	*xylA*	A19535	184	19 890	Maat et al. (1992)
Aspergillus niger	XynNB	*xynNB*	D38071	188	20 069	Kinoshita et al. (1995)
Aspergillus niger	Xyn5	*XYN5*	U39784	195	21 143	Unpublished
Aspergillus oryzae	XynG1	*xynG1*	AB003085	189	20 589	Kimura et al. (1998)
Aspergillus tubigensis	XYLA	*xlnA*	L26988	184	19 837	de Graaff et al. (1994)
Aspergillus tubigensis	Xylanase B		A39368	207	22 240	Patent WO 9414965-A (1994)
Aureobasidium pullulans	XynA	*xynA*	U10298	187	20 074	Li and Ljungdahl (1994)
Chaetomium gracile	CgXA	*cgxA*	D49850	189	20 149	Yoshino et al. (1995)
Chaetomium gracile	CgXB	*cgxB*	D49851	211	22 525	Yoshino et al. (1995)
Claviceps purpurea	Xylanase	*xyl1*	Y16969	197	21 460	Giesbert et al. (1998)
Cochliobolus carbonum	Xyl1	*XYL1*	L13596	191	20 856	Apel et al. (1993)
Cochliobolus carbonum	Xyl2	*XYL2*	U58915	191	21 199	Apel-Birkhold and Walton (1996)
Cochliobolus carbonum	Xyl3	*XYL3*	U58916	183	19 858	Apel-Birkhold and Walton (1996)
Cochliobolus sativus	Xylanase	*xyl2*	AJ004802	212	23 658	Emami and Hack (2001)
Cryptococcus sp. S-2	Xyn-CS2	*xyn-CS2*	D63382	184	20 209	Iefuji et al. (1996)
Humicola insolens	Xyl1	*xyl1*	X76047	208	23 814	Dalboge and Heldt-Hansen (1994)
Magnaporthe grisea	XYN22	*xyn22*	L37529	194	21 427	Wu et al. (1995)
Neocallimastix frontalis	Xylanase 2	*XYN2*	S48865	212	23 933	Unpublished
Neocallimastix frontalis	XYN3 (domA)[b]	*xyn3*	X82266	223	24 394	Durand et al. (1996)
Neocallimastix frontalis	XYN3 (domB)[b]	*xyn3*	X82266	223	24 532	Durand et al. (1996)
Neocallimastix patriciarum	XYLA dom1)[b]	*xynA*	X65526	226	24 941	Gilbert et al. (1992)
Neocallimastix patriciarum	XYLA (dom2)[b]	*xynA*	X65526	225	24 770	Gilbert et al. (1992)
Orpinomyces strain PC-2	XynA[b]	*xynA*	U57819	225	24 810	Li et al. (1997)
Paecilomyces varioti	PVX		P81536[c]	194	21 365	Kumar et al. (2000)
Penicillium sp. 40	XynA	*xynA*	AB035540	190	20 713	Kimura et al. (2000)
Penicillium purpurogenum	XynB	*xynB*	Z50050	183	19 371	Díaz et al. (1997)
Pichia stipitis	Xylanase A[b]	*xynA*	AF151379	232	26 291	Unpublished
Piromyces sp. (inactive)	XYLA[b]	*xynA*	X91858	234	25 558	Fanutti et al. (1995)
Piromyces sp. (active)	XYLA[b]	*xynA*	X91858	222	24 803	Fanutti et al. (1995)
Schizophyllum commune	Xylanase A	*xynA*	P35809[c]	197	20 965	Oku et al. (1993)
Thermomyces lanuginosus	XynA	*xynA*	U35436	206	22 614	Schlacher et al. (1996)
Trichoderma harzianum E58	20 kD Xylanase		P48793[c]	190	20 690	Yaguchi et al. (1992b)
Trichoderma reesei	XYNI	*xyn1*	X69574	178	19 035	Törrönen et al. (1992)
Trichoderma reesei	XYNII	*xyn2*	X69573	190	20 829	Törrönen et al. (1992)
Trichoderma reesei	Xyn2	*XYN2*	U24191	190	20 731	La Grange et al. (1996)
Trichoderma viride	Xylanase IIA		A44594	190	20 759	Yaguchi et al. (1992a)
Trichoderma viride	Xylanase IIB		A44595	190	20 743	Unpublished
Protozoa						
Polyplastron multivesiculatum	XYN A	*xynA*	AJ009828	219	25 192	Unpublished
Polyplastron multivesiculatum	Xylanase	*polyX*	AB011274	175	19 394	Unpublished
Insect						
Phaedon cochleariae	Xylanase		Y17908	200	22 070	Unpublished

[a] Number of amino acid residues.
[b] Multidomain protein.
[c] SWISS-PROT entry.

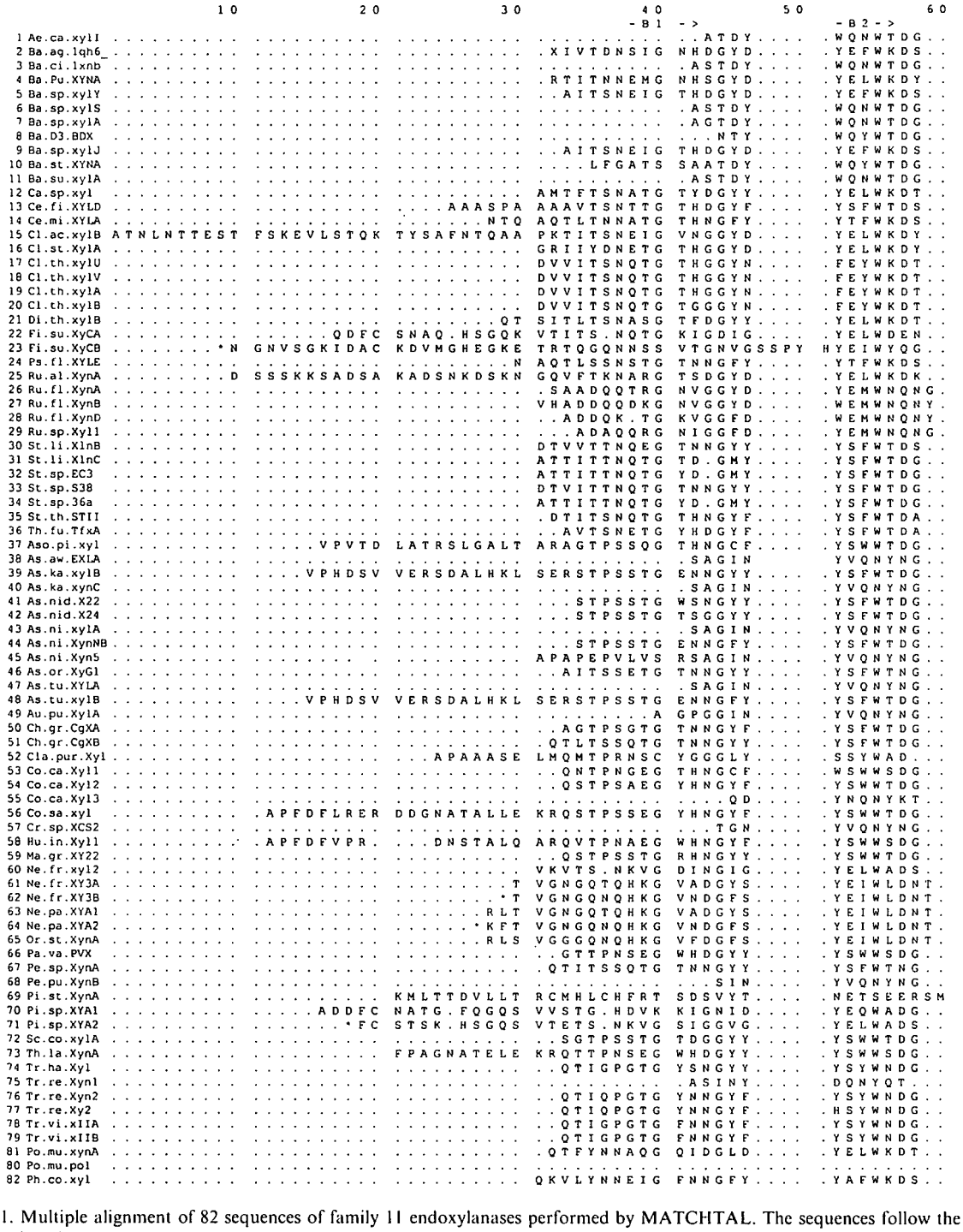


Fig. 1. Multiple alignment of 82 sequences of family 11 endoxylanases performed by MATCHTAL. The sequences follow the same order given in Table 1. On top of the alignment, secondary structure elements observed in the crystal structures are provided. Boxes (Predicted Structural Conserved Regions) obtained by MATCHBOX are shaded. The two catalytic Glu residues (E169 and E289 according to the numbering of the multiple alignment) are in boldface. Residue 100 (N or D, depending on the pH optimum, see text) is underlined. The putative start and/or end of the catalytic domain sequences belonging to multiple domain xylanases is indicated by a star. For abbreviations see Table 1.

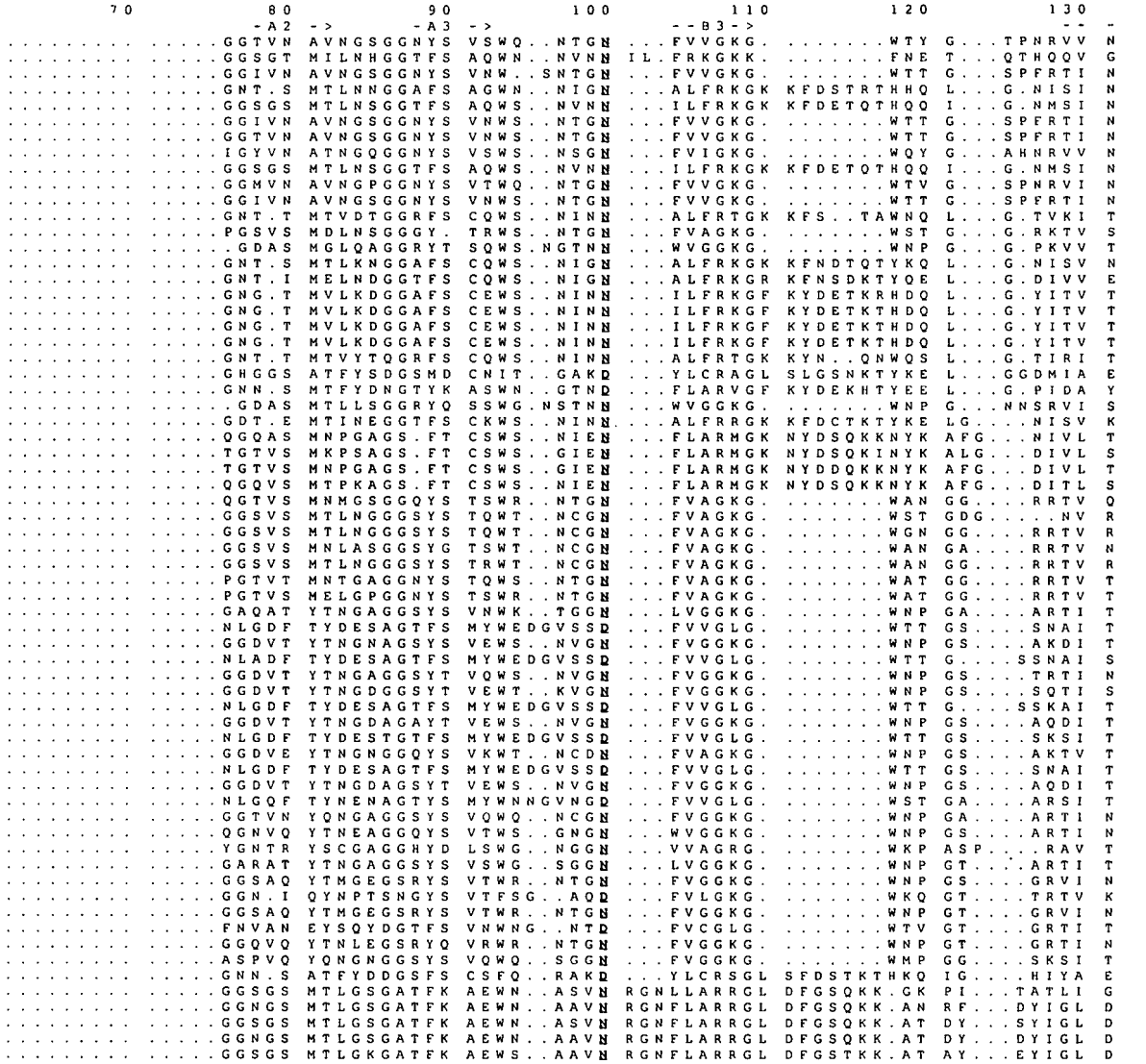


Fig. 1. (*Continued*)

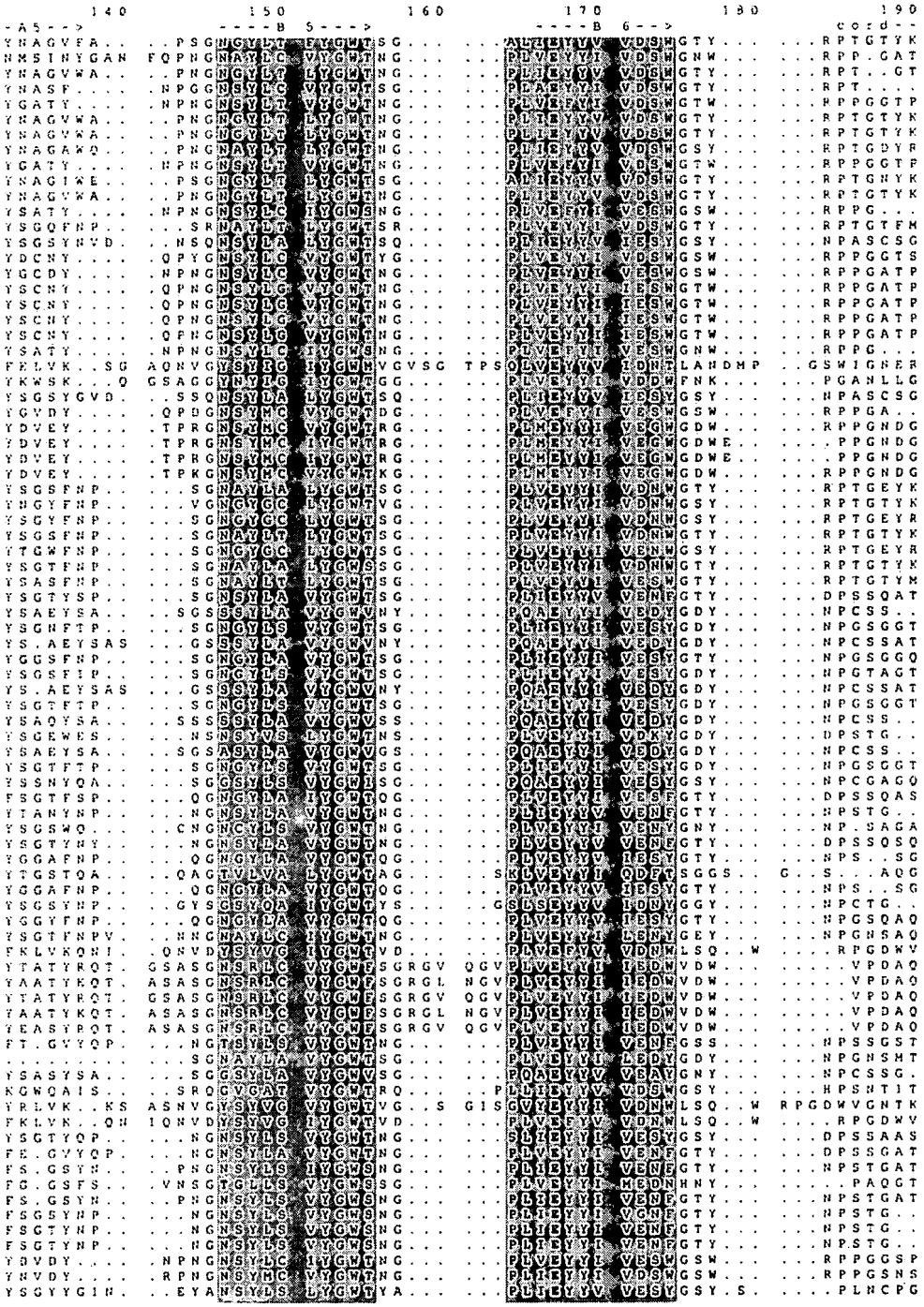


Fig. 1. (*Continued*)

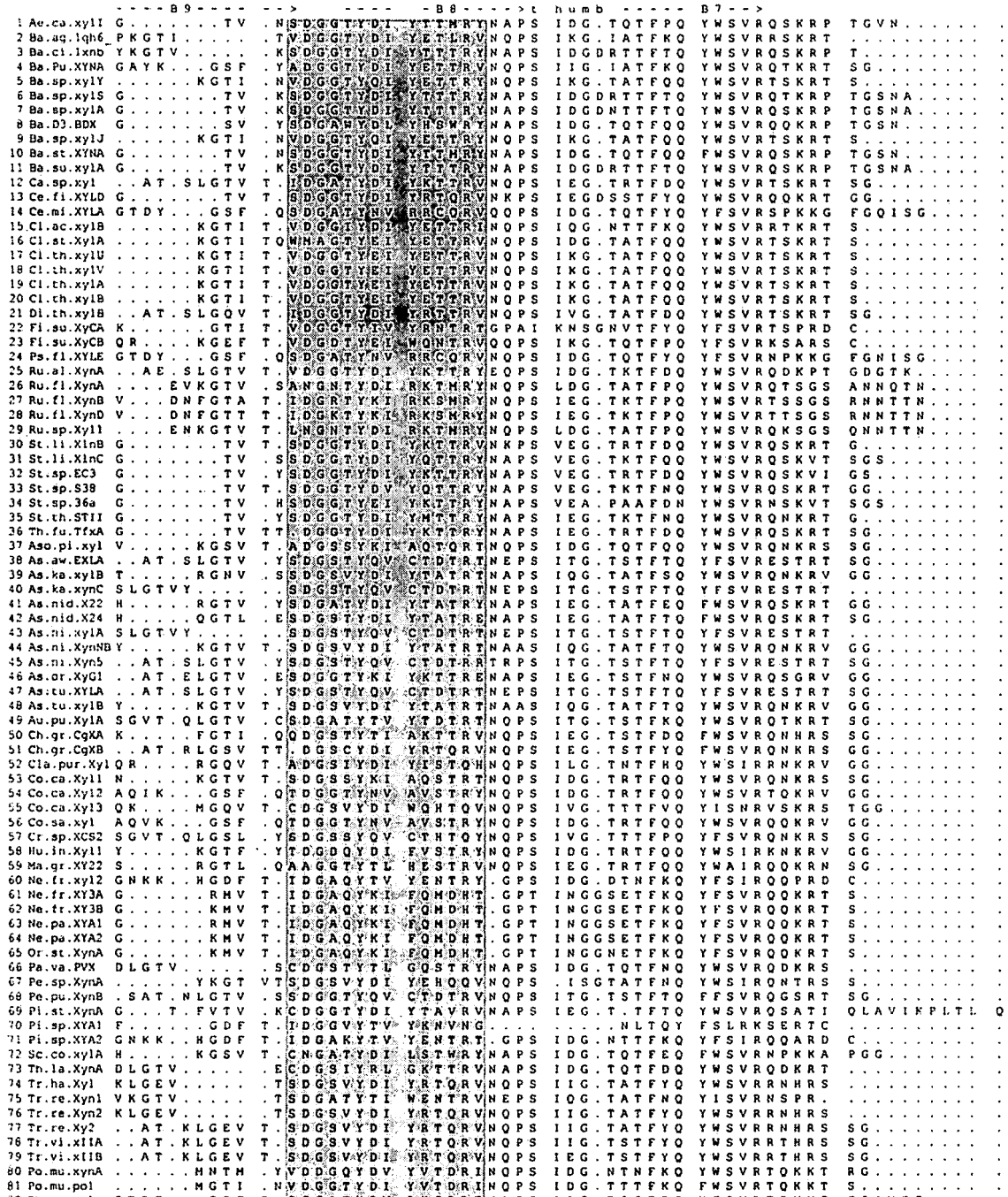


Fig. 1. (*Continued*)

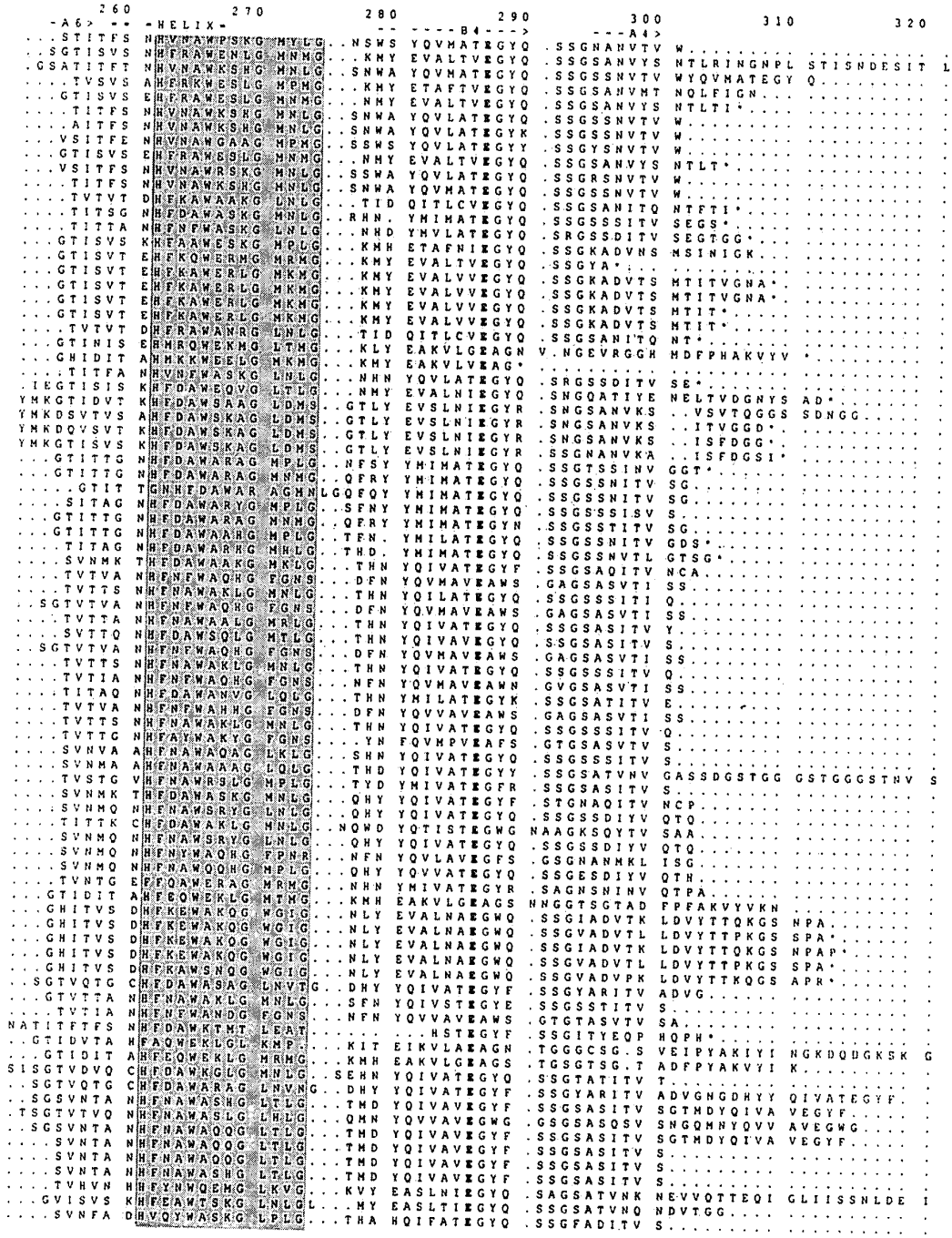


Fig. 1. (*Continued*)

Table 2
Isoelectric point, pH and T° optimum of family 11 xylanases of known sequence

Organism	Xylanase	pI	pH optimum	Optimum T°	Reference
Bacteria					
Aeromonas caviae	Xylanase I	7.1	7.0	55	Kubata et al. (1992)
Bacillus agradhaerens	xyl 11	8.8	n.d.	n.d.	Sabini et al. (1999)
Bacillus pumilus	XYNA	n.d.	6.5	45–60	Panbangred et al. (1983)
Bacillus sp. D3	Xylanase	7.7	6.0	75	Harris et al. (1997)
Bacillus sp. 41 M1	Xylanase J	5.5	9.0	50	Nakamura et al. (1992)
Bacillus subtilis	Xylanase A	8.9	n.d	n.d.	Paice et al. (1986)
Caldicellulosiruptor sp. Rt69B.1	Xylanase	n.d.	5.5	70	Morris et al. (1999)
Clostridium acetobutylicum	Xylanase B	8.5	5.5–6	60	Lee et al. (1987)
Clostridium stercorarium	XynA	4.5	7.0	75	Sakka et al. (1994)
Clostridium thermocellum	XynA	n.d.	6.5	65	Hayashi et al. (1999)
Dictroglonus thermophilum	Xylanase B	n.d	6.5	85	Morris et al. (1998)
Fibrobacter succinogenes	XynC	6.2	6.5	n.d.	Paradis et al. (1993)
Ruminococcus flavefaciens	XYLA	5.0	5.5	50	Flint et al. (1991) García-Campayo et al. (1993)
Streptomyces lividans	XlnB	8.4	6.5	55	Kluepfel et al. (1990)
Streptomyces lividans	XlnC	>10.25	6.0	57	Kluepfel et al. (1992)
Streptomyces sp. EC3	Xylanase	9.1	n.d.	n.d.	Mazy-Servais et al. (1996)
Streptomyces sp. S38	Xyl1	9.8	6.0–6.5	55–60	Georis et al. (2000)
Streptomyces thermoviolaceus	STX-II	8.0	7.0	60	Tsujibo et al. (1992)
Thermomonospora fusca	TfxA	10	7.0	n.d.	Irwin et al. (1994)
Fungi					
Aspergillus awamori	EXLA	3.7	n.d.	n.d.	Hessing et al. (1994)
Aspergillus kawachii	XynC	3.5	2.0	50	Ito et al. (1992)
Aspergillus nidulans	X22	6.4	5.5	62	Fernández-Espinar et al. (1993)
Aspergillus nidulans	X24	3.5	5.5	52	Fernández-Espinar et al. (1996)
Aspergillus niger	Xyl A (or I)	3.7	3.0	n.d.	Maat et al. (1992) Krengel and Dijkstra (1996)
Aspergillus niger	XynNB	n.d.	5.0	n.d.	Kinoshita et al. (1995)
Aspergillus tubigensis	XYLA	3.6	n.d.	n.d.	de Graaff et al. (1994)
Aureobasidium pullulans	XynA (APX II)	9.4	4.8	54	Li et al. (1993)
Cochliobolus carbonum	Xyl1	>9.3	4.0–8.0	45	Holden and Walton (1992)
Cochliobolus carbonum	Xyl2 and 3	>9.3	5.0	n.d.	Holden and Walton (1992)
Cryptococcus sp. S-2	Xyn-CS2	7.4	2.0	40	Iefuji et al. (1996)
Magnaporthe grisea	XYN22	9.7	n.d.	n.d.	Wu et al. (1995)
Paecilomyces varioti	PVX	3.9	5.5–7.0	65	Krishnamurthy and Vithayathil (1989)
Penicillium sp. 40	XynA	4.7	2.0	50	Kimura et al. (2000)
Penicillium purpurogenum	XynB	5.9	3.5	50	Belancic et al. (1995)
Schizophyllum commune	Xylanase A	4.5	5.0	50	Jurasek and Paice (1988)
Thermomyces lanuginosus	XynA	4.1	6.5	65	Gomes et al. (1993) Schlacher et al. (1996)
Trichoderma harzianum E58	20 kD xylanase	9.4	5.0	50	Wong and Saddler (1992)
Trichoderma reesei	XYNI	5.2	3.5–4	n.d.	Törrönen et al. (1992)
Trichoderma reesei	XYNII	9.0	4.5–5.5	n.d.	Törrönen et al. (1992)
Trichoderma viride	Xylanase IIA	9.3	5.0	53	Wong and Saddler (1992)

n.d.: not determined.

Table 3
Family 11 endoxylanases of known three-dimensional structure

Organism	Protein	PDB code	References
Bacteria			
Bacillus agaradhaerens	Xylanase	1qh6	Sabini et al. (1999)
Bacillus circulans	XLNA	1xnb	Wakarchuk et al. (1994a)
Bacillus sp. D3	Xylanase[a]		Harris et al. (1997)
Dictyoglomus thermophilum	XynB[a]	1fsj	McCarthy et al. (2000)
Fungi			
Aspergillus kawachii	XynC	1bk1	Fushinobu et al. (1998)
Aspergillus niger	XylI or A	1ukr	Krengel and Dijkstra (1996)
Paecilomyces varioti	Xylanase[a]	1pvx	Kumar et al. (2000)
Thermomyces lanuginosus	XynA[a]	1yna	Gruber et al. (1998)
Trichoderma harzianum E58	Xylanase	1xnd	Campbell et al. (1993)
Trichoderma reesei	XYNI	1xyn	Törrönen and Rouvinen (1995)
Trichoderma reesei	XYNII	1xyp	Törrönen et al. (1994)

[a] Thermophilic protein.

package (MSI, San Diego) or the FSSP program (Holm and Sander, 1996). In addition, the use of CLUSTAL to globally align the regions inserted between these anchor points allows to highlight several key residues in less conserved regions. In short, the alignment obtained on the basis of sequence similarity also reflects structural similarities among the endoxylanases.

The alignment is shown in Fig. 1. The residues are numbered on top; the numbering exceeds the length of any individual sequence since it includes all gaps assigned by the alignment. The secondary structure elements are numbered as shown in Fig. 2.

Four boxes (shaded in Fig. 1) are found along the whole set of 82 sequences and correspond to the PSCRs. Interestingly, the sequence alignment provides a pattern of aligned segments and gaps consistent with the observed secondary structure elements. Indeed, the lowest sequence homology is found in the regions between beta strands (residues 94–103 between A3 and B3; residues 111–130 between B3 and A5; residues 238–254 between B7 and A6). The region corresponding to the thumb, (between strands B7 and B8) is, however, very well conserved along all the sequences, except for *Piromyces* sp. XYLA. This last enzyme is reported to be the inactive form of the protein, in contrast to *Piromyces* sp. XYLB, the active form, that also contains the conserved segment defining the thumb region. The cord region (between B6 and B9) is also conserved. The PSCR's, in particular, correspond to secondary structure elements: box 1 to B5, box 2 to B6, box 3 to B8 and box 4 to the helix.

The amino terminal regions of the aligned sequences show no similarities in the first 30 or so residues. One sequence is particularly long in this region (*C. acetobutylicum*). One of the sequences (*P. multivesiculatum*, AB011274) starts only at residue 81, and lacks strands B1, B2 and part of A2, suggesting that B1 and B2 may not be necessary for enzymatic activity. Residue 40 is a G in 60 sequences, and is missing in 15. It belongs to B1 in the structure of *T. reesei* XYN II (Törrönen et al., 1994). Its importance is unclear, since B1 is missing totally or in part in the indicated 15 sequences.

The segment corresponding to B2 shows good similarity among the sequences. Position 52 is in most cases an aromatic residue: 63 sequences have a Y, while 14 show a W or a F. At position 55, W predominates (70 sequences), while 62 sequences have a D in position 57. Between B2 and A2, two long insertions (probably forming a loop) are found in the endoxylanases from *Penicillium* sp. and *Pichia stipitis*. 50 G's are observed in position 76 and 50 in position 77. Strand A2 shows low similarity except at position 81, where 24 Y and

35 M are found. In the loop linking A2 and A3, 61 G's are observed in position 86 and 67 in position 87, but not necessarily together in the same sequence. In A3, a highly conserved F or Y (79 sequences) is found in position 89 and 75 W in 93; in this last case, the few replacements are F or Y; only one non-conserved replacement (H) is observed in *P. stipitis*.

Residue 100 is in all cases a D (27) or N (65); as we will see below, this residue is related to the pH optimum of the enzyme. At the end of B3, G109 is present in 79 sequences (it is replaced by a K in *Bacillus agaradhaerens*) and allows a special twist to the chain (Törrönen and Rouvinen, 1997).

A5 is not present in *P. stipitis*, which shows a big gap between 110 and 144; in this sequence, B3 and B5 are probably linked by a short loop. B5 corresponds to the first box detected by MATCHBOX, indicating a highly conserved stretch of the sequence, particularly YGW (152–154), present in all 82 enzymes. Y152 is strongly hydrogen bonded to E167; it seems to be of great functional importance since its mutation to F (in *Bacillus circulans* XLNA) leads to a totally inactive enzyme (Wakarchuk et al., 1994a). At position 150 21 Cys are present, the highest number found at a particular location in all the sequences (see below).

The next box corresponds to B6. It is preceded by a highly conserved Pro at 164 (present in 74 sequences) and Leu 165 (74 sequences). E167 is common to all sequences and corresponds to one of the catalytic residues (the nucleophile); it is followed by two highly conserved Y's (75 in position 168, replaced in the rest of the enzymes by F;

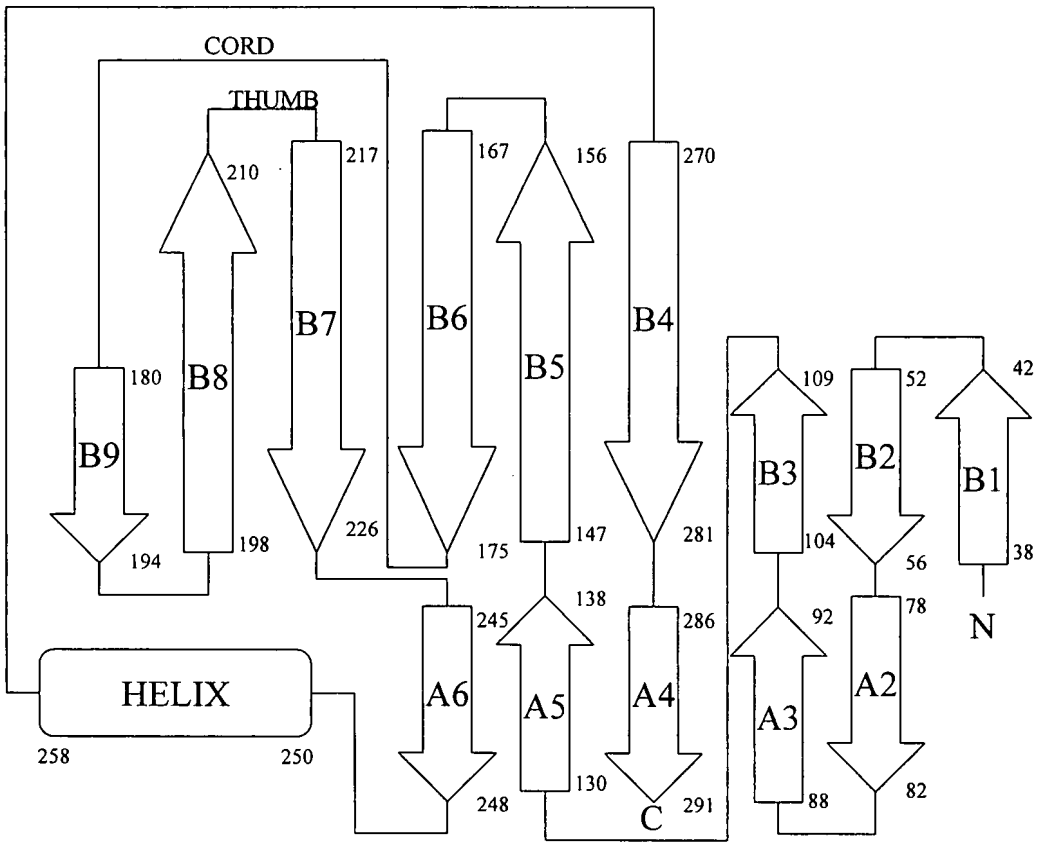


Fig. 2. Topology diagram of the family 11 endoxylanases. The residue numbering refers to that of Fig. 1. The figure is adapted from Törrönen et al. (1994).

and common to all in 169) and by two aliphatic hydrophobic residues in positions 170 and 171. Tyr 169 links both catalytic glutamates, acting as a charge stabilizing residue (Törrönen and Rouvinen, 1997), and its mutation to F in XLNA of *B. circulans* leads to a large decrease in activity (Wakarchuk et al., 1994a).

The 'cord' shows low similarity, except for a Pro (found in 70 sequences and missing in 7) in position 185. B9 is a stretch of very low similarity and varies considerably in length, but it is followed by the third high similarity box, corresponding mainly to B8. Between B9 and B8, two highly conserved residues are apparent: D204 (in 76 sequences) and G205 (in 80 sequences); this last residue is considered important in hairpin formation (Törrönen and Rouvinen, 1997). Y208 is present in all 82 sequences, R215 in 72 and N217 in 69 (this last residue is missing in 8 sequences).

The 'thumb' is well conserved; it has a consensus sequence PSIXG where X is almost any residue and the others show few and mainly homologous replacements. Pro 219 gives a twist to the strand at the beginning of the thumb. B7 is also well conserved; residue 227 is a T in 77 sequences and 228 is F in 81 cases; a consensus sequence QYWSVR can be proposed for residues 230–235 (R is replaced by K in one instance and the other residues show mainly conservative replacements).

The loop connecting B7 and A6 differs in length; the longest being in the enzyme from *P. stipitis*, which also shows the longest loop between B2 and A2. No significant similarities are apparent in A6, which is immediately followed by the helix.

The last box of high similarity comprises the carboxyl end of the α-helix and the following 6–7 residues. H262 is present in 79 sequences, W263 in 80 and G270 in 79.

B4 does not show high similarities, except for E287 (boldface in Fig. 1), a residue present in all sequences, which corresponds to the acid-base glutamate participating in catalysis, and which is followed by a highly conserved G (70 sequences). The last well defined secondary structure is A4. A highly conserved consensus sequence SSGS precedes and starts this strand.

The carboxyl terminus of the sequences is highly variable in length. The longest is found in *B. agaradhaerens* and *P. multivesiculatum*, stretching 25 and 22 residues, respectively, beyond the end of A4.

3.2. Acidophily and thermostability

A number of family 11 endoxylanases have been reported to be acidophilic. It has been postulated that for xylanases that function optimally under acidic conditions, residues spatially adjacent to the acid/base catalytic glutamate influence the pH optimum (Törrönen and Rouvinen, 1995). In particular, the substitution of N100 (underlined in Fig. 1) by D shifts the pH optimum from 5.7 to 4.6, as has been demonstrated by mutational, kinetic, and structural studies of N100D in *B. circulans* XYLA (Joshi et al., 2000). This agrees with the mutational analysis of xylanase C of *A. kawachii*, in which the single substitution of Asp 100 to Asn at this key position dramatically elevates its pH optimum from 2 to 5 (Fushinobu et al., 1998). Structural studies of xylanase A (or I) from *A. niger* led to a similar conclusion. In the crystal structure of this enzyme of low pH optimum (Krengel and Dijkstra, 1996), Asp 100 is assigned a critical role.

An examination of the enzymes of known sequence for which the pH optimum has been determined (Table 2), shows that this correlation holds true with only one exception. The enzymes showing a pH optimum below 5 have D at position 100, while N is present in those with pH optima of 5 or more. The exception is XynC from *F. succinogenes*, an anaerobic bacterium from the rumen. This enzyme shows a pH optimum of 6.5 but has a D at position 100. This enzyme has two catalytic domains, and the pH optimum reported in Table 2 corresponds to the native enzyme; the separate domains have a pH optimum of 6.0 (Zhu et al., 1994).

In conclusion, there is a strong correlation in that the residue hydrogen bonded to the general acid/base catalyst at position 100 is asparagine in the so-called 'alkaline' xylanases, whereas it is aspartic acid in those with a more acidic pH optimum.

Thermostability is an important issue in the properties of endoxylanases, due to their biotechnological applications, particularly in cellulose biobleaching. Of the enzymes listed in Table 2,

seven of bacterial origin (*Bacillus* D3, *Caldicellulosiruptor* sp., *Clostridium stercorarium*, *Clostridium thermocellum*, *Dictyoglomus thermophilum*, *Streptomyces thermoviolaceus* and *Thermomonospora fusca*) and two from fungi (*Paecilomyces varioti* and *Thermomyces lanuginosus*) are considered thermophilic, based on their optimal temperature and stability at high temperature. Table 3 shows that the three-dimensional structure of four of these enzymes has been determined.

Thermophilicity and thermostability may be explained by a variety of factors and structural parameters (Kumar et al., 2000). Of those, the importance of S–S bridges and aromatic 'sticky patches' can be analyzed by sequence alignment. Additional structural features potentially involved in thermal stability, such as salt bridges, aromatic interactions and entropic effects have been postulated in family 11 xylanases (Georis et al., 2000).

Cysteine residues are not very common in these enzymes. Of the total number of residues in the 82 sequences listed in Table 1, only 117 are Cys, a 0.7%. Twenty-seven of the sequences have no Cys and 14 have only one; therefore, at least half of the total possesses no disulfide bridges. The three-dimensional structure shows that an S–S bridge is found in xylanase A from *A. niger* connecting Cys 186 to Cys 211, thus attaching the cord to the large beta-sheet (B8) (Krengel and Dijkstra, 1996). These two half-cystine residues are conserved in a few other fungal family 11 xylanases (*A. kawachii*, *Aspergillus awamori*, *A. niger* Xyn5, *Aspergillus tubigensis*, *Cryptococcus* sp. and *Penicillium purpurogenum*). The xylanases from the bacteria *Cellvibrio mixtus* and *Pseudomonas fluorescens* and of the insect *Phaedon cochleariae* have Cys at positions 188 and 213, which may be forming a similar bridge. The presence of this disulfide bridge may influence the stability of these proteins, although it does not give (at least to the enzymes analyzed so far) a thermophilic character.

Of the seven thermophilic xylanases listed, three have no cysteines, and the three-dimensional structure of the *Dictyoglomus thermophilus* enzyme shows no S–S bonds, although its sequence has 3 Cys. On the other hand, *T. lanuginosus* and *P. varioti* xylanases, both thermophilic, do have an S–S bridge linking Cys 203 (located in B9) and Cys 261 (in the α helix). *Cochliobolus carbonum* Xyl3 and *Schizophyllum commune* XylA also possess only two Cys and in the same positions; the T° optimum of the former has not been reported, while the latter has a value of only 50 °C (Table 2). All these results suggest that S–S bridges are unlikely to be of importance in the thermophilicity of family 11 xylanases.

As pointed out by Turunen et al. (2001), thermophilicity (activity at high temperature) and thermostability do not necessarily depend on the same structural factors. They introduced a disulfide bridge plus other minor mutations in *T. reesei* XynII (C203-C261) significantly increasing the thermostability without affecting the temperature optimum. Wakarchuk et al. (1994b) have introduced an S–S bridge at the same position in *B. circulans* xylanase, obtaining similar results. Thus, the introduction of S–S bridges to enzymes may affect both properties differently.

Harris et al. (1997) propose that the *Bacillus* D3 xylanase (lacking S–S bridges) is stabilized through 'sticky patches' between pairs of molecules through the interaction of surface aromatic residues. These residues are Tyr 53, 78, 120, 201, 290, 295 and Trp 207 and 214. When the alignment is analyzed, it shows that Tyr 78, 120 and Trp 207 are unique to this sequence (no aromatic residues are present in those positions in the remaining 81 sequences), while Trp 214 is shared by only 1 sequence (not thermophilic) and no other aromatic residues are found in this location. The remaining four residues show no clear alignment pattern. Therefore, it can be concluded that the 'sticky patch' pattern described for the *Bacillus* D3 xylanase is very unique, and it is unlikely to play an important role in stabilizing the other thermophilic family 11 xylanases of known sequence.

Shibuya et al. (2000) using random gene shuffling between a mesophilic (*S. lividans* XlnB) and a thermophilic (*T. fusca* TfxA) enzyme have shown the importance of the amino terminal segment of the protein in temperature stability. However, if the first 50 residues (our numbering) of the thermophilic enzymes in this study are aligned, no

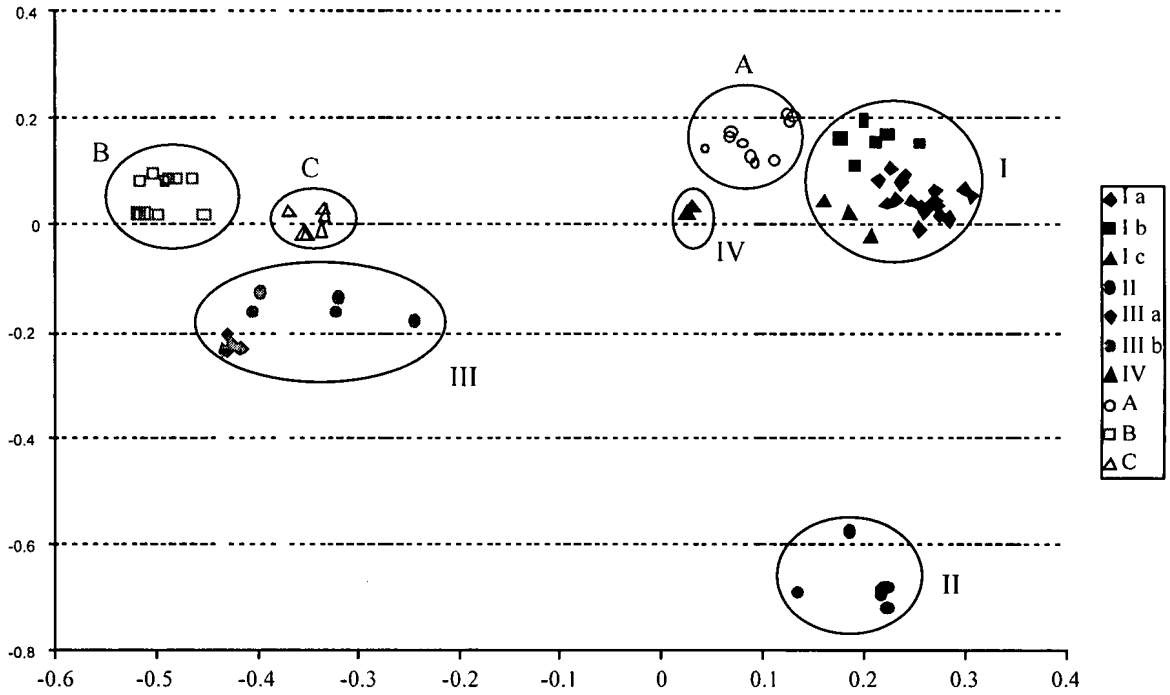


Fig. 3. Results from the factor analysis. Graphical representation of the sequences in the plane of factors 2 and 3. Each point corresponds to a sequence, the position on the plot being associated to a percentage of the total variability within the data set. Distances separating each position in the plot are directly related to the sequence homology. Groups containing mainly sequences of fungi (I, II, III, IV) are marked by filled symbols.

clear similarity pattern is observed, indicating that, again, the proposed thermostabilizing factor may be unique to the *T. fusca* enzyme.

3.3. Classification of the family 11 xylanases

A factor analysis (Explore subprogram of MATCHBOX) was performed in order to define groups and subgroups among the 82 sequences under study. Sequences have been represented in a three-dimensional space, each factor being associated to a percentage of the total variability between the sequences. The first factor is generally trivial, and the grouping of the sequences was performed in the plane of factors 2 and 3 (Depiereux and Feytmans, 1992). Graphical representation of the sequences in the plane of factors 2 and 3 (Fig. 3) was obtained directly from the program MATCHBOX and represented via Excel. The distances separating each position in the plot are directly related to the sequence homology, thus allowing a classification. Sequences close in the plane of factors 2 and 3 were grouped/clustered, leading to the classification presented in Table 4. A first clustering, based on the distance between coordinates in the plane of factors 2 and 3, led to the definition of a total of seven groups. Those groups were labeled (A, B, C, I, II, III, IV) *a posteriori*, on the basis of the origin (fungal or bacterial) of the sequences belonging to them. The two larger groups (I and III), were further divided into subgroups on the basis of the distance between coordinates in the plane of factors 2 and 3 within the group. Group B, although containing many sequences (13) was not divided into subgroups as the distribution of points is narrow (factor 2 comprised between -0.453 and -0.518; factor 3 comprised between 0.015 and 0.087) in comparison to the distribution of, for example, group III ($-0.431 <$ factor $2 < -0.240$; $-0.236 <$ factor $3 < -0.161$).

This method avoids affecting the sequence classification by misalignments occurring in a previous

step, and represents the sequences on a plane taking into account Euclidian distance rather than hierarchical merging. The graphical representation of the eigenvectors (sequence coordinates) in the plane of factors 2 and 3 (Fig. 3) allows to classify and group the sequences by eye, since the distances separating each position in the plot are directly related to differences in the sequences. The classification obtained has been validated by a clustering (Ward method) performed on the sequence coordinates.

Interestingly, the groups obtained from the factor analysis are highly homogeneous, meaning that they mostly contain only sequences from either bacteria or fungi. It is also noteworthy that some bacterial sequences are more homologous to sequences of fungal endoxylanases than to sequences of other bacteria (see for example group I).

In this work a robust classification methodology based on factor analysis, without previous sequence alignment and without reference to any phylogenetic inference is applied for the first time to classify a large set of sequences. Interestingly, the classification of xylanases presented in Table 4 compares favorably with previous classifications based on smaller sequence datasets. In particular, our results are in good agreement with the classification presented by Georis et al. (1999) deduced from a phylogenetic tree analysis on 63 sequences. According to their study, family 11 endoxylanases were subdivided into six main groups: three for

Table 4
Classification of family 11 xylanases based on a factor analysis

Ia	Tr.re.XYNI	II	As.ni.XylA	*A*	*Ce.fi.XYLD*
Ia	Tr.re.XYNII	II	Pe.pu.XynB	*A*	*Th.fu.TfxA*
Ia	Tr.vi.xIIA	II	As.tu.XYLA	*A*	*St.sp.S38*
Ia	Tr.vi.xIIB	II	As.ni.Xyn5	*A*	*St.th.STII*
Ia	Th.la.XynA	II	As.ka.XynC	*A*	*St.li.XlnC*
Ia	Pe.sp.XynA	II	As.aw.EXLA	*A*	*Ba.ci.1xnb*
Ia	Pa.va.1pvx	II	Au.pu.XylA	*A*	*St.sp.EC3*
Ia	Ch.gr.CgXB			*A*	*St.li.XlnB*
Ia	Aso.pi.xyl	IIIa	Ne.fr.XY3A	*A*	*Ba.su.xylA*
Ia	As.or.XyG1	IIIa	Ne.fr.XY3B	*A*	*Ba.sp.xylA*
Ia	As.nid.X24	IIIa	Ne.pa.XYA1		
Ia	As.ka.xylB	IIIa	Ne.pa.XYA2	*C*	*Di.th.xylB*
Ia	Tr.ha.1xnd	IIIa	Orpin.XynA	*C*	*Ru.fl.XynB*
Ia	As.nid.X22			*C*	*Ru.fl.XynD*
Ia	Ma.gr.XY22	IIIb	Fi.su.XyCA	*C*	*Ca.sp.xyl*
Ia	Co.ca.Xyl1	IIIb	Fi.su.XyCB	*C*	*Ru.sp.Xyl1*
Ia	Ch.gr.CgXA	IIIb	Pi.sp.XYA1	*C*	*Ru.fl.XynA*
Ia	As.tu.xylB	IIIb	Pi.sp.XYA2		
Ia	As.ni.XynNB	IIIb	Ne.fr.xyl2	*B*	*Ba.ag.1qh6*
				B	*Ru.al.XynA*
Ib	Co.ca.Xyl2	IV	Co.ca.Xyl3	*B*	Po.mu.POX
Ib	Cla.pu.Xyl	IV	Pi.st.XynA	*B*	Po.mu.xylA
Ib	Co.sa.xyl			*B*	*Cl.ac.xylB*
Ib	Sc.co.xylA			*B*	*Ba.sp.xylJ*
Ib	Hu.in.Xyl1			*B*	*Ba.sp.xylY*
Ib	Ba.D3.BDX			*B*	*Cl.th.xylV*
				B	*Cl.st.XylA*
Ic	*Ce.mi.*XYLA			*B*	*Cl.th.xylA*
Ic	Ps.fl.XYLE			*B*	*Cl.th.xylB*
Ic	Ph.co.xyl			*B*	*Cl.th.xylU*
				B	*Ba.pu.XYNA*

Sequence of bacterial enzymes are in italics and enzymes produced by protozooa or insects are underlined. For abbreviations see Table 1 and Fig. 1.

fungi, two for Gram positive bacteria and one for Gram negative bacteria.

A similar subdivision is found here: groups I, II and III contain mainly fungal enzymes. The enzymes in groups I and II are mostly the 20 kDa enzymes from *Ascomyceta* and *Basidiomyceta*. Most enzymes of group I exhibit a basic pI. Those in group II show an acidic pI. Enzymes of group III are mainly produced by anaerobic fungi. Like in the classification proposed by Georis et al. (1999) two enzymes produced by *F. succinogenes* (XyCA and XyCB), an anaerobic Gram-negative bacterium, also fall in this group.

In addition to those three groups associated mainly to fungal enzymes, a fourth group that was not present in previous classifications emerges from the present factor analysis. It only contains two enzymes. One is produced by *C. carbonum* (Xyl3) and is clearly distinct from Xyl1 and Xyl2 produced by the same fungus. In the classification proposed by Georis et al. (1999), this enzyme is isolated in the unrooted phylogenetic tree. The second enzyme in group IV is produced by *P. stipitis* (XynA); this enzyme has not been included in any previous classification.

Bacterial enzymes are mainly divided into three groups (A, B, C). Group A contains mainly enzymes produced by members of the *Actinomycetaceae* and the *Bacillaceae* families, strictly aerobic Gram-positive bacteria. Groups B and C are more closely related and contain mainly enzymes from anaerobic Gram-positive bacteria, such as those from *Clostridium* or *Ruminococcus*, which usually live in the rumen. Two bacterial enzymes, XylE and XylA xylanases from *P. fluorescens* and *C. mixtus*, respectively, strictly aerobic Gram-negative bacteria, are found in subgroup Ic. In terms of sequence similarities, those two Gram-negative bacterial enzymes more closely resemble xylanases produced by fungi (group I) than other Gram-positive bacterial enzymes (e.g. groups A, B, C). They were also classified in a distinct group by Georis et al. (1999).

3.4. Biotechnological significance

Xylanases are finding an increasing number of applications, both alone and in combination with other enzymes. Among them are cellulose pulp biobleaching (Buchert et al., 1994), bread-making (Courtin et al., 1999) and saccharification of lignocellulosic biomass (Lee, 1997). These applications require enzymes capable of operating under specific and often unnatural conditions. Parameters of particular interest are thermostability and pH optimum. Biobleaching, for instance, requires thermostable and alkali-stable enzymes. Family 11 xylanases may be of particular interest in biobleaching due to their smaller size; a fact which may facilitate penetration in the cellulose fiber network.

Optimizing enzyme properties for a particular application can be achieved by random and site-directed mutagenesis. Arase et al. (1993) have achieved a significant stabilization of *Bacillus pumilus* XynA by random mutagenesis. Four heat-resistant mutants were isolated, and the stabilizing mutations were found to be clustered in the N-terminal region. In three of the four mutants, a mutation of G92 to S or to D was observed. Residue 92 is located in the A3 strand, and it is always a polar or charged residue except in *B. pumilus* and *Penicillium* sp. where it is a G (Fig. 1); the introduction of an S or a D at that position may allow the formation of a stabilizing hydrogen bond.

An example of the use of site-directed mutagenesis is given by Turunen et al. (2001). They have introduced a disulfide bridge by means of the mutations S203C–N261C, thus increasing significantly the half-life of the enzyme at 65°. As shown in Fig. 1, the thermophilic xylanase from *T. lanuginosus* possesses this bridge.

It has proved difficult to manipulate the properties of an enzyme in a predictable manner (Törrönen and Rouvinen, 1997). However, for the design and protein engineering of endoxylanases with properties suitable for their different applications, a good knowledge of its sequence and structure is a necessary condition. The information and analysis presented in this publication should be useful for this purpose.

Acknowledgements

This work was financed in part by grants from DIPUC (Dirección de Investigación y Postgrado,

Pontificia Universidad Católica de Chile) and FONDECYT grants 1960241, 3950015 and 8990004 (Líneas Complementarias).

References

Apel, P.C., Panaccione, D.G., Holden, F.R., Walton, J.D., 1993. Cloning and targeted gene disruption of XYL1, a β-1,4-xylanase gene from the maize pathogen *Cochliobolus carbonum*. Molec. Plant Microbe Interact. 6, 467–473.

Apel-Birkhold, P.C., Walton, J.D., 1996. Cloning, disruption and expression of two endo-β 1,4-xylanase genes, XYL2 and XYL3 from *Cochliobolus carbonum*. Appl. Environ. Microbiol. 62, 4129–4135.

Arase, A., Yomo, T., Urabe, I., Hata, Y., Katsube, Y., Okada, H., 1993. Stabilization of xylanase by random mutagenesis. FEBS Lett. 316, 123–127.

Belancic, A., Scarpa, J., Peirano, A., Díaz, R., Steiner, J., Eyzaguirre, J., 1995. *Penicillium purpurogenum* produces several xylanases: purification and properties of two of the enzymes. J. Biotechnol. 41, 71–79.

Bertrand, L., Vertommen, D., Depiereux, E., Hue, L., Rider, M.H., Feytmans, E., 1997. Modelling the 2-kinase domain of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase on adenylate kinase. Biochem. J. 321, 615–621 erratum appears in Biochem. J. 323 (1997) 864.

Bertrand, L., Vertommen, D., Freeman, P.M., Wouters, J., Depiereux, E., Di Pietro, A., Hue, L., Rider, M.H., 1998. Mutagenesis of the fructose-6-phosphate-binding site in the 2-kinase domain of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. Eur. J. Biochem. 254, 490–496.

Biely, P., 1985. Microbial xylanolytic systems. Trends Biotechnol. 3, 286–290.

Biely, P., Vrsanská, M., Tenkanen, M., Kluepfel, D., 1997. Endo-β-1,4-xylanase families: differences in catalytic properties. J. Biotechnol. 57, 151–166.

Briffeuil, P., Baudoux, G., Lambert, C., De Bolle, X., Vinals, C., Feytmans, E., Depiereux, E., 1998. Comparative analysis of seven multiple protein sequence alignment servers: clues to enhance reliability of predictions. Bioinformatics 14, 357–366.

Buchert, J., Tenkanen, M., Kantelinen, A., Viikari, L., 1994. Application of xylanases in the pulp and paper industry. Bioresource Technol. 50, 65–72.

Campbell, R., Rose, D., Wakarchuk, W., To, R., Sung, W., Yaguchi, M., 1993. A comparison of the structures of the 20 kD xylanases from *Trichoderma harzianum* and *Bacillus circulans*. In: Suouminen, P., Reinikainen, T. (Eds.), Proceedings of the Second TRICEL Symposium on *Trichoderma reesei* Cellulases and other Hydrolases. Espoo, Finland, pp. 63–72.

Cho, S., Choi, Y., 1995. Nucleotide sequence analysis of an endo-xylanase gene (xynA) from *Bacillus stearothermophilus*. J. Microbiol. Biotechnol. 5, 117–124.

Courtin, C.M., Roelants, A., Delcour, J.A., 1999. Fractionation-reconstitution experiments provide insight into the role of endoxylanases in bread-making. J. Agric. Food Chem. 47, 1870–1877.

Dalboge, H., Heldt-Hansen, H.P., 1994. A novel method for efficient expression cloning of fungal enzyme genes. Mol. Gen. Genet. 243, 253–260.

De Bolle, X., Vinals, C., Prozzi, D., Paquet, J.-Y., Leplae, R., Depiereux, E., Vandenhaute, J., Feytmans, E., 1995. Identification of residues potentially involved in the interactions between subunits in yeast alcohol dehydrogenases. Eur. J. Biochem. 231, 214–219.

de Fays, K., Tibor, A., Lambert, C., Vinals, C., Denoel, P., De Bolle, X., Wouters, J., Letesson, J.J., Depiereux, E., 1999. Structure and function prediction of the *Brucella abortus* P39 protein by comparative modeling with marginal sequence similarities. Prot. Eng. 12, 217–223.

de Graaff, L.H., van den Broeck, H.C., van Ooijen, A.J.J., Visser, J., 1994. Regulation of the xylanase-encoding xlnA gene of *Aspergillus tubigensis*. Mol. Microbiol. 12, 479–490.

Depiereux, E., Feytmans, E., 1991. Simultaneous and multivariate alignment of protein sequences. Correspondence between physico-chemical profiles and structurally conserved regions (SCR's). Prot. Eng. 4, 603–613.

Depiereux, E., Feytmans, E., 1992. MATCH-BOX: a fundamentally new algorithm for the simultaneous alignment of several protein sequences. Comp. App. Biosci. 8, 501–509.

Depiereux, E., Baudoux, G., Briffeuil, P., Reginste, I., De Bolle, X., Vinals, C., Feytmans, E., 1997. Match-Box_ server: a multiple sequence alignment tool placing emphasis on reliability. Comp. App. Biosci. 13, 249–256.

Díaz, R., Sapag, A., Peirano, A., Steiner, J., Eyzaguirre, J., 1997. Cloning, sequencing and expression of the cDNA of endoxylanase B from *Penicillium purpurogenum*. Gene 187, 247–251.

Durand, R., Rascle, C., Fevre, M., 1996. Molecular characterization of *xyn* 3, a member of the endoxylanase multigene family of the rumen anaerobic fungus *Neocallimastix frontalis*. Curr. Genet. 30, 531–540.

Emami, K., Hack, E., 2001. Characterization of a xylanase gene from *Cochliobolus sativus* and its expression. Mycol. Res. 105, 352–359.

Fanutti, C., Ponyi, T., Black, G.W., Hazlewood, G.P., Gilbert, H.J., 1995. The conserved noncatalytic 40-residue sequence in cellulases and hemicellulases from anaerobic fungi functions as a protein docking domain. J. Biol. Chem. 270, 29314–29322.

Fernández-Espinar, M.T., Piñaga, F., Sanz, P., Ramón, D., Vallés, S., 1993. Purification and characterization of a neutral endoxylanase from *Aspergillus nidulans*. FEMS Microbiol. Lett. 113, 223–228.

Fernández-Espinar, M.T., Vallés, S., Piñaga, F., Pérez-González, J.A., Ramón, D., 1996. Construction of an *Aspergillus nidulans* multicopy transformant for the *xlnB* gene and its use in purifying the minor X_{24} xylanase. Appl. Microbiol. Biotechnol. 45, 338–341.

Flint, H.J., McPherson, C.A., Martin, J., 1991. Expression of two xylanase genes from the rumen cellulolytic bacterium *Ruminococcus flavefaciens* cloned in pUC13. J. Gen. Microbiol. 137, 123–129.

Flint, H.J., Martin, J., McPherson, C.A., Daniel, A.S., Zhang, J.-X., 1993. A bifunctional enzyme with separate xylanase and β-(1,3-1,4) glucanase domains, encoded by the *xynD* gene of *Ruminococcus flavefaciens*. J. Bacteriol. 175, 2943–2951.

Fukusaki, E., Panbangred, W., Shinmyo, A., Okada, H., 1984. The complete nucleotide sequence of the xylanase gene (xynA) of *Bacillus pumilus*. FEBS Lett. 171, 197–201.

Fushinobu, S., Ito, K., Konno, M., Wakagi, T., Matsuzawa, H., 1998. Crystallographic and mutational analyses of an extremely acidophilic and acid-stable xylanase: Biased distribution of acidic residues and importance of Asp 37 for catalysis at low pH. Prot. Eng. 11, 1121–1128.

García-Campayo, V., McCrae, S., Zhang, J.-X., Flint, H.J., Wood, T.M., 1993. Mode of action, kinetic properties and physicochemical characterization of two different domains of a bifunctional (1→4)-β-D-xylanase from *Ruminococcus flavefaciens* expressed separately in *Escherichia coli*. Biochem. J. 296, 235–243.

Georis, J., Giannotta, F., Lamotte-Brasseur, J., Devreese, B., van Beeumen, J., Granier, B., Frère, J.M., 1999. Sequence, overproduction and purification of the family 11 endo-beta-1,4, xylanase encoded by the xyl1 gene of *Streptomyces* sp. S38. Gene 237, 123–133.

Georis, J., de Lemos-Esteves, F., Lamotte-Brasseur, J., Bougnet, V., Devreese, B., Giannotta, F., Granier, B., Frère, J.M., 2000. An additional aromatic interaction improves the thermostability and thermophilicity of a mesophilic family 11 xylanase: structural basis and molecular study. Prot. Sci. 9, 466–475.

Giesbert, S., Leppping, H.-B., Tenberge, K.B., Tudzynski, P., 1998. The xylanolytic system of *Claviceps purpurea*: cytological evidence for secretion of xylanases in infected rye tissue and molecular characterization of two xylanase genes. Phytopathology 88, 1020–1030.

Gilbert, H.J., Hazlewood, G.P., Laurie, J.I., Orpin, C.G., Xue, G.P., 1992. Homologous catalytic domains in a rumen fungal xylanase: evidence for gene duplication and prokaryotic origin. Mol. Microbiol. 6, 2065–2072.

Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C. Jr., Warren, R.A.J., 1991. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. 55, 303–315.

Gomes, J., Gomes, I., Kreiner, W., Esterbauer, H., Sinner, M., Steiner, W., 1993. Production of high level of cellulase-free and thermostable xylanase by a wild strain of *Thermomyces lanuginosus* using beechwood xylan. J. Biotechnol. 30, 283–297.

Gruber, K., Klintschar, G., Hayn, M., Schlacher, A., Steiner, W., Kratky, C., 1998. Thermophilic xylanase from *Thermomyces lanuginosus*: high-resolution X-ray structure and modeling studies. Biochemistry 37, 13475–13485.

Harris, G.W., Jenkins, J.A., Connerton, I., Pickersgill, R.W., 1996. Refined crystal structure of the catalytic domain of xylanase A from *Pseudomonas fluorescens* at 1.8 Å resolution. Acta Crystallog. Sec. D 52, 393–401.

Harris, G.W., Pickersgill, R.W., Connerton, I., Debeire, P., Touzel, J.-P., Breton, C., Pérez, S., 1997. Structural basis of the properties of an industrially relevant thermophilic xylanase. Proteins 29, 77–86.

Hayashi, H., Takehara, M., Hattori, T., Kimura, T., Karita, S., Sakka, K., Ohmiya, K., 1999. Nucleotide sequences of two contiguous and highly homologous xylanase genes *xynA* and *xynB* and characterization of XynA from *Clostridium thermocellum*. Appl. Microbiol. Biotech. 51, 348–357.

Henrissat, B., Davies, G., 1997. Structural and sequence-based classification of glycoside hydrolases. Curr. Opin. Struct. Biol. 7, 637–644.

Hessing, J.G.M., van Rotterdam, C.O., Verbakel, J.M.A., Roza, M., Maat, J., van Gorcom, R.F.M., van den Hondel, C.A.M.J.J., 1994. Isolation and characterization of a 1,4-β-endoxylanase gene of *Aspergillus awamori*. Curr. Genet. 26, 228–232.

Higgins, D.G., Bleasby, A.J., Fuchs, R., 1992. CLUSTAL V: improved software for multiple sequence alignment. CABIOS 8, 189–191.

Himmel, M.E., Karplus, P.A., Sakon, J., Adney, W.S., Baker, J.O., Thomas, S.R., 1997. Polysaccharide hydrolase folds. Diversity of structure and convergence of function. Appl. Biochem. Biotechnol. 63–65, 315–325.

Holden, R., Walton, J.D., 1992. Xylanases from the fungal maize pathogen *Cochliobolus carbonum*. Physiol. Mol. Plant Pathol. 50, 39–47.

Holm, L., Sander, C., 1996. Mapping the protein universe. Science 273, 595–603.

Iefuji, H., Chino, M., Kato, M., Iimura, Y., 1996. Acid xylanase from yeast *Cryptococcus* sp. S-2: purification, characterization, cloning and sequencing. Biosci. Biotechnol. Biochem. 60, 1331–1338.

Irwin, D., Jung, E.D., Wilson, D.B., 1994. Characterization and sequence of a *Thermomonospora fusca* xylanase. Appl. Environ. Microbiol. 60, 763–770.

Ito, K., Ogasawara, H., Sugimoto, T., Ishikawa, T., 1992. Purification and properties of acid stable xylanases from *Aspergillus kawachii*. Biosci. Biotechnol. Biochem. 56, 547–550.

Ito, K., Iwashita, K., Iwano, K., 1992a. Cloning and sequencing of the *xynC* genes encoding acid xylanase of *Aspergillus kawachii*. Biosci. Biotechnol. Biochem. 56, 1338–1340.

Jeong, K.J., Lee, P.C., Park, I.Y., Kim, M.S., Kim, S.C., 1998. Molecular cloning and characterization of an endoxylanase gene of *Bacillus* sp. in *Escherichia coli*. Enzyme Microb. Technol. 22, 599–605.

Joseleau, J.P., Comptat, J., Ruel, K., 1992. Chemical structure of xylans and their interaction in the plant cell walls. In: Visser, J., Beldman, G., Kusters-Van Someren, M.A., Voragen, A.G.J. (Eds.), Xylans and Xylanases. Elsevier, Amsterdam, pp. 1–15.

Joshi, M.D., Sidhu, G., Pot, I., Brayer, G.D., Withers, S.G., McIntosh, L.P., 2000. Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. J. Mol. Biol. 299, 255–279.

Jurasek, L., Paice, M.G., 1988. Xylanase A of *Schizophyllum commune*. Meth. Enzymol. 160, 659–662.

Kimura, T., Kitamoto, N., Kito, Y., Karita, S., Sakka, K., Ohmiya, K., 1998. Molecular cloning of xylanase gene *xynG1* from *Aspergillus oryzae* KBN 616, a shoyu koji mold, and analysis of its expression. J. Ferment. Bioeng. 85, 10–16.

Kimura, T., Ito, J., Kawano, A., Makino, T., Kondo, H., Karita, S., Sakka, K., Ohmiya, K., 2000. Purification, characterization and molecular cloning of acidophilic xylanase from *Penicillium* sp. 40. Biosci. Biotechnol. Biochem. 64, 1230–1237.

Kinoshita, K., Takano, M., Koseki, T., Ito, K., Iwano, K., 1995. Cloning of the *xynNB* gene encoding xylanase B from *Aspergillus niger* and its expression in *Aspergillus kawachii*. J. Ferment. Bioeng. 79, 422–428.

Kluepfel, D., Vats-Mehta, S., Aumont, F., Shareck, F., Morosoli, R., 1990. Purification and characterization of a new xylanase (xylanase B) produced by *Streptomyces lividans* 66. Biochem. J. 267, 45–50.

Kluepfel, D., Daigneault, N., Morosoli, R., Shareck, F., 1992. Purification and characterization of a new xylanase (xylanase C) produced by *Streptomyces lividans*. Appl. Microbiol. Biotechnol. 36, 626–631.

Krengel, U., Dijkstra, B.W., 1996. Three-dimensional structure of endo-1,4-β-xylanase I from *Aspergillus niger*: molecular basis for its low pH optimum. J. Mol. Biol. 263, 70–78.

Krishnamurthy, S., Vithayathil, P.J., 1989. Purification and characterization of endo-1,4-β-xylanase from *Paecilomyces varioti* Bainier. J. Ferment. Bioeng. 67, 77–82.

Kubata, B.K., Horitsu, H., Kawai, K., Takamizawa, K., Suzuki, T., 1992. Xylanase I of *Aeromonas caviae* ME-1 isolated from the intestine of a herbivorous insect (*Samia cynthia pryeri*). Biosci. Biotechnol. Biochem. 56, 1463–1464.

Kubata, B.K., Suzuki, T., Ito, Y., Naito, H., Kawai, K., Takamizawa, K., Horitsu, H., 1997. Cloning and expression of xylanase I gene (*xynA*) of *Aeromonas caviae* ME-1 in *Escherichia coli*. J. Ferment. Bioeng. 83, 292–295.

Kumar, P.R., Eswaramoorthy, S., Vithayathil, P.J., Viswamitra, M.A., 2000. The tertiary structure at 1.59 Å resolution and the proposed amino acid sequence of a family-11 xylanase from the thermophilic fungus *Paecilomyces varioti* Bainier. J. Mol. Biol. 295, 581–593.

La Grange, D.C., Pretorius, I.S., van Zyl, W.H., 1996. Expression of a *Trichoderma reesei* β-xylanase gene (XYN2) in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 62, 1036–1044.

Lee, J., 1997. Biological conversion of lignocellulosic biomass to ethanol. J. Biotechnol. 56, 1–24.

Lee, S.F., Forsberg, C.W., Rattray, J.B., 1987. Purification and characterization of two endoxylanases from *Clostridium acetobutylicum* ATCC 824. Appl. Environ. Microbiol. 53, 644–650.

Li, X.L., Zhang, Z.Q., Dean, J.F., Eriksson, K.E., Ljungdahl, L.G., 1993. Purification and characterization of a new xylanase (APX-II) from the fungus *Aureobasidium pullulans* Y-2311-1. Appl. Environ. Microbiol. 59, 3212–3218.

Li, X.L., Ljungdahl, L.G., 1994. Cloning, sequencing, and regulation of a xylanase gene from the fungus *Aureobasidium pullulans* Y-2311-1. Appl. Environ. Microbiol. 60, 3160–3166 published erratum in: Appl. Environ. Microbiol. 60 (1994) 4647.

Li, X.L., Chen, H.Z., Lungdahl, L.G., 1997. Monocentric and polycentric anaerobic fungi produce structurally related cellulases and xylanases. Appl. Environ. Microbiol. 63, 628–635.

Lübeck, P.D., Paulin, L., Degefu, Y., Lübeck, M., Alekhina, I., Bulat, S.A., Collinge, D.B., 1997. PCR cloning, DNA sequencing and phylogenetic analysis of a xylanase gene from the phytopathogenic fungus *Ascochyta pisi Lib.* Physiol. Mol. Plant. Pathol. 51, 377–389.

Maat, J., Roza, M., Verbakel, J., Stam, H., Santos da Silva, M.J., Bosse, M., Egmond, M.R., Hagemans, M.L.D., van Gorcom, R.F.M., Hessing, J.G.M., van den Hondel, C.A.M.J.J., van Rotterdam, C.O., 1992. Xylanases and their application in bakery. In: Visser, J., Beldman, G., Kusters-Van Someren, M.A., Voragen, A.G.J. (Eds.), Xylans and Xylanases. Elsevier, Amsterdam, pp. 349–360.

Mazy-Servais, C., Moreau, A., Gerard, C., Dusart, J., 1996. Cloning and nucleotide sequence of a xylanase-encoding gene from *Streptomyces* sp. strain EC3. DNA Sequence 6, 147–158.

McCarthy, A.A., Morris, D.D., Bergquist, P.L., Baker, E.N., 2000. Structure of XynB, a highly thermostable β-1,4-xylanase from *Dictyoglomus thermophilum* Rt46B.1, at 1.8 Å resolution. Acta Cryst. D 56, 1366–1375.

Millward-Sadler, S.J., Poole, D.M., Henrissat, B., Hazlewood, G.P., Clarke, J.H., Gilbert, H.J., 1994. Evidence for a general role for high-affinity non-catalytic cellulose binding domains in microbial plant cell wall hydrolases. Mol. Microbiol. 11, 375–382.

Millward-Sadler, S.J., Davidson, K., Hazlewood, G.P., Black, G.W., Gilbert, H.J., Clarke, J.H., 1995. Novel cellulose-binding domains, NodB homologues and conserved modular architecture in xylanases from the aerobic soil bacteria *Pseudomonas fluorescens* subsp. *cellulosa* and *Cellvibrio mixtus*. Biochem. J. 312, 39–48.

Morris, D.O., Gibbs, M.D., Chin, C.W.J., Koh, M.H., Wong, K.K.Y., Allison, R.W., Nelson, P.J., Bergquist, P.L., 1998. Cloning of the *xynB* gene from *Dictyoglomus thermophilum* Rt 46 B.1 and action of the gene product on kraft pulp. Appl. Environ. Microbiol. 64, 1759–1765.

Morris, D.O., Gibbs, M.D., Ford, M., Thomas, J., Bergquist, P.L., 1999. Family 10 and 11 xylanase genes from *Caldicellulosiruptor* sp. strain Rt69B.1. Extremophiles 3, 103–111.

Nagashima, M., Okumoto, Y., Okanishi, M., 1989. Nucleotide sequence of the gene of extracellular xylanase in *Streptomyces* sp. No.36 a and construction of secretion vectors using xylanase gene. Trends Actinomycetol., 91–96.

Nakai, R., Wakabayashi, K., Asano, T., Aono, R., Horikoshi, K., Nakamura, S., 1994. Nucleotide sequence and mutational analysis of the gene encoding the novel alkaline xylanase from alkaliphilic *Bacillus* sp. strain 41M-1. Nucleic Acids Symp. Ser. 31, 235–236.

Nakamura, S., Aono, R., Wakabayashi, K., Horikoshi, K., 1992. Alkaline xylanase produced by newly isolated alkaliphilic *Bacillus* sp. In: Visser, J., Beldman, G., Kusters-Van Someren, M.A., Voragen, A.G.J. (Eds.), Xylans and Xylanases. Elsevier, Amsterdam, pp. 443–446.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Prot. Eng. 10, 1–6.

Oku, T., Roy, C., Watson, D.C., Wakarchuk, W., Campbell, R., Yaguchi, M., Jurasek, L., Paice, M.G., 1993. Amino acid sequence and thermostability of xylanase A from *Schizophyllum commune*. FEBS Lett. 334, 296–300.

Paice, M.G., Bourbonnais, R., Desrochers, M., Jurasek, L., Yaguchi, M., 1986. A xylanase gene from *Bacillus subtilis*: nucleotide sequence and comparison with *Bacillus pumilus* gene. Arch. Microbiol. 144, 201–206.

Panbangred, W., Shinmyo, A., Kinoshita, S., Okada, H., 1983. Purification and properties of endoxylanase produced by *Bacillus pumilus*. Agric. Biol. Chem. 47, 957–963.

Paradis, F.W., Zhu, H., Krell, P.J., Phillips, J.P., Forsberg, C.W., 1993. The *xynC* gene from *Fibrobacter succinogenes* S85 codes for a xylanase with two similar catalytic domains. J. Bacteriol. 175, 7666–7672.

Perez-González, J.A., de Graaff, L.H., Visser, J., Ramón, D., 1996. Molecular cloning and expression in *Saccharomyces cerevisiae* of two *Aspergillus nidulans* xylanase genes. Appl. Environ. Microbiol. 62, 2179–2182.

Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jorgensen, P.L., Schülein, M., Dupont, C., Davies, G.J., Wilson, K.S., 1999. Catalysis and specificity in enzymatic glycoside hydrolysis: a ^{2,5}B conformation for the glycosyl-enzyme intermediate revealed by the structure of the *Bacillus agaradhaerens* family 11 xylanase. Chem. Biol. 6, 483–492.

Sakka, K., Kojima, Y., Kondo, T., Karita, S.-I., Ohmiya, K., Shimada, K., 1993. Nucleotide sequence of the *Clostridium stercorarium xynA* gene encoding xylanase A: identification of catalytic and cellulose binding domains. Biosci. Biotechnol. Biochem. 57, 273–277.

Sakka, K., Kojima, Y., Kondo, T., Karita, S.-I., Shimada, K., Ohmiya, K., 1994. Purification and characterization of xylanase A from *Clostridium stercorarium* F-9 and a recombinant *E. coli*. Biosci. Biotechnol. Biochem. 58, 1496–1505.

Schlacher, A., Holzmann, K., Hayn, M., Steiner, W., Schwab, H., 1996. Cloning and characterization of the gene from the thermostable xylanase XynA from *Thermomyces lanuginosus*. J. Biotechnol. 49, 211–218.

Shareck, F., Roy, C., Yaguchi, M., Morosoli, R., Kluepfel, D., 1991. Sequences of three genes specifying xylanases in *Streptomyces lividans*. Gene 107, 75–82.

Shibuya, H., Kaneko, S., Hayashi, K., 2000. Enhancement of the thermostability and hydrolytic activity of xylanase by random gene shuffling. Biochem. J. 349, 651–656.

Sunna, A., Antranikian, G., 1997. Xylanolytic enzymes from fungi and bacteria. Crit. Rev. Biotechnol. 17, 39–67.

Törrönen, A., Mach, R.L., Messner, R., González, R., Kalkkinen, N., Harkki, A., Kubicek, C.P., 1992. The two major xylanases from *Trichoderma reesei*: characterization of both enzymes and genes. Bio./Technol. 10, 1461–1465.

Törrönen, A., Harkki, A., Rouvinen, J., 1994. Three-dimensional structure of endo-1,4-β-xylanase II from *Trichoderma reesei*: two conformational states in the active site. EMBO J. 13, 2493–2501.

Törrönen, A., Rouvinen, J., 1995. Structural comparison of two major endo-1,4-xylanases from *Trichoderma reesei*. Biochemistry 34, 847–856.

Törrönen, A., Rouvinen, J., 1997. Structural and functional properties of low molecular weight endo-1,4-β-xylanases. J. Biotechnol. 57, 137–149.

Tsujibo, H., Miyamoto, K., Kuda, T., Minami, K., Sakamoto, T., Hasegawa, T., Inamori, Y., 1992. Purification, properties and partial amino acid sequences of thermostable xylanase from *Streptomyces thermoviolaceus* OPC-520. Appl. Environ. Microbiol. 58, 371–375.

Tsujibo, H., Ohtsuki, T., Iio, T., Yanazaki, I., Miyamoto, K., Sugiyama, M., Inamori, Y., 1997. Cloning and sequence analysis of genes encoding xylanases and acetyl xylan esterase from *Streptomyces thermoviolaceus* OPC-520. Appl. Environ. Microbiol. 63, 661–664.

Turunen, O., Etuaho, K., Fenel, F., Vehmaanperä, J., Wu, X., Rouvinen, J., Leisola, M., 2001. A combination of weakly stabilizing mutations with a disulfide bridge in the α-helix region of *Trichoderma reesei* endo-1,4-β-xylanase II increases the thermal stability through synergism. J. Biotechnol. 88, 37–46.

Vinals, C., De Bolle, X., Depiereux, E., Feytmans, E., 1995. Knowledge-based modelling of the D-lactate dehydrogenase three-dimensional structure. Prot. Struc. Funct. Genet. 21, 307–318.

Wakarchuk, W.W., Campbell, R.L., Sung, W.L., Davoodi, J., Yaguchi, M., 1994a. Mutational and crystallographic analysis of the active site residues of the *Bacillus circulans* xylanase. Protein Sci. 3, 467–475.

Wakarchuk, W.W., Sung, W.L., Campbell, R.L., Cunningham, A., Watson, D.C., Yaguchi, M., 1994b. Thermostabilization of the *Bacillus circulans* xylanase by the introduction of disulfide bonds. Prot. Eng. 7, 1379–1386.

Wong, K.K.Y., Saddler, J.N., 1992. *Trichoderma* xylanases, their properties and application. Crit. Rev. Biotechnol. 12, 413–435.

Wu, S.-C., Kaufman, S., Darvill, A.G., Albersheim, P., 1995. Purification, cloning and characterization of two xylanases from *Magnaporthe grisea*, the rice blast fungus. Mol. Plant–Microbe Interact. 8, 506–514.

Yaguchi, M., Roy, C., Ujiie, M., Watson, D.C., Wakarchuk, W., 1992a. Amino acid sequence of the low-molecular weight xylanase from *Trichoderma viride*. In: Visser, J., Beldman, G., Kusters-Van Someren, M.A., Voragen, A.G.J. (Eds.), Xylans and Xylanases. Elsevier, Amsterdam, pp. 149–154.

Yaguchi, M., Roy, C., Watson, D.C., Rollin, F., Tan, L.U.L., Senior, D.J., Saddler, N., 1992b. The amino acid sequence of the 20KD xylanase from *Trichoderma harzianum* E58. In: Visser, J., Beldman, G., Kusters-Van Someren, M.A., Voragen, A.G.J. (Eds.), Xylans and Xylanases. Elsevier, Amsterdam, pp. 435–438.

Yang, R.C.A., Mackenzie, C.R., Narang, S.A., 1988. Nucleotide sequence of a *Bacillus circulans* xylanase gene. Nucleic Acids Res. 16, 7187.

Yoshino, S., Oishi, M., Moriyama, R., Kato, M., Tsukagoshi, N., 1995. Two family G xylanase genes from *Chaetomium gracile* and their expression in *Aspergillus nidulans*. Curr. Genet. 29, 73–80.

Yu, J.H., Park, Y.S., Yum, D.Y., Kim, J.M., Kong, I.S., Bai, D.H., 1993. Nucleotide sequence and analysis of a xylanase gene (*xynS*) from alkali-tolerant *Bacillus* sp. YA-14 and comparison with other xylanases. J. Microbiol. Biotechnol. 3, 139–145.

Zappe, H., Jones, W.A., Woods, D.R., 1990. Nucleotide sequence of a *Clostridium acetobutylicum* P262 xylanase gene (*xynB*). Nucleic Acids Res. 18, 2179.

Zhang, J.-X., Flint, H.J., 1992. A bifunctional xylanase encoded by the *xynA* gene of the rumen cellulolytic bacterium *Ruminococcus flavefaciens* 17 comprises two dissimilar domains linked by an asparagine/glutamine-rich sequence. Mol. Microbiol. 6, 1013–1023.

Zhang, J.-X., Martin, J., Flint, H.J., 1994. Identification of non-catalytic conserved regions in xylanases encoded by the *xynB* and *xynD* genes of the cellulolytic rumen anaerobe *Ruminococcus flavefaciens*. Mol. Gen. Genet. 245, 260–264.

Zhu, H., Paradis, F.W., Krell, P.J., Phillips, J.P., Forsberg, C.W., 1994. Enzymatic specificities and modes of action of the two catalytic domains of the XynC xylanase from *Fibrobacter succinogenes* S85. J. Bacteriol. 176, 3885–3894.

Eur. J. Biochem. **270**, 1399–1412 (2003) © FEBS 2003 doi:10.1046/j.1432-1033.2003.03496.x

Three-dimensional structures of thermophilic β-1,4-xylanases from *Chaetomium thermophilum* and *Nonomuraea flexuosa*

Comparison of twelve xylanases in relation to their thermal stability

Nina Hakulinen[1], Ossi Turunen[2], Janne Jänis[1], Matti Leisola[2] and Juha Rouvinen[1]

[1]*Department of Chemistry, University of Joensuu, Finland;* [2]*Helsinki University of Technology, Finland*

The crystal structures of thermophilic xylanases from *Chaetomium thermophilum* and *Nonomuraea flexuosa* were determined at 1.75 and 2.1 Å resolution, respectively. Both enzymes have the overall fold typical to family 11 xylanases with two highly twisted β-sheets forming a large cleft. The comparison of 12 crystal structures of family 11 xylanases from both mesophilic and thermophilic organisms showed that the structures of different xylanases are very similar. The sequence identity differences correlated well with the structural differences. Several minor modifications appeared to be responsible for the increased thermal stability of family 11 xylanases: (a) higher Thr : Ser ratio (b) increased number of charged residues, especially Arg, resulting in enhanced polar interactions, and (c) improved stabilization of secondary structures involved the higher number of residues in the β-strands and stabilization of the α-helix region. Some members of family 11 xylanases have a unique strategy to improve their stability, such as a higher number of ion pairs or aromatic residues on protein surface, a more compact structure, a tighter packing, and insertions at some regions resulting in enhanced interactions.

Keywords: xylanase; glycoside hydrolases; family 11; thermostability.

Xylanases (EC 3.2.1.8) are glycoside hydrolases that catalyze the hydrolysis of internal β-1,4 bonds of xylan, the major hemicellulose component of the plant cell wall. The enzymatic hydrolysis of xylan has potential economical and environment-friendly applications. Xylanases can be used in bleaching of pulp to reduce the use of toxic chlorine-containing chemicals [1] or to improve the quality of animal feed [2]. In addition, there are applications in the food and beverage industry [3]. Therefore, attention is focused on discovery of new xylanases or improvement of existing ones in order to meet the requirements of industry such as stability and activity at high temperature and extreme pH.

The xylanases that have been structurally characterized to date can be classified into the glycoside hydrolase families 10 and 11, corresponding to former families F and G, respectively [4]. Family 10 enzymes have an $(\alpha/\beta)_8$ barrel fold with a molecular mass of approximately 35 kDa. Family 11 xylanases are somewhat smaller, approximately 20 kDa, and their fold contains an α-helix and two β-sheets packed against each other, forming a so-called β-sandwich. Due to the industrial applications of xylanase, both xylanase families are well studied. In this paper, we focus on xylanases in family 11.

To date, the crystal structures of family 11 xylanases are available from several organisms: *Trichoderma harzianum* [5], *Bacillus circulans* [5–7], *Trichoderma reesei* [8,9], *Aspergillus niger* [10], *Thermomyces lanuginosus* [11], *Aspergillus kawachii* [12], *Bacillus agaradhaerens* [13], *Paecilomyces varioti* [14], and *Dictyoglomus thermophilum* [15]. Three of these, *T. lanuginosus*, *P. varioti*, and *D. thermophilum* are from thermophilic organisms. In addition, a low-resolution structure has been reported for thermostable *Bacillus* D3 [16] but no PDB coordinates are available. Very recently, the structures of two new xylanases from *Streptomyces* sp. S38 [17] and *Bacillus subtilis* B230 [18] have also been solved. A disulphide bridge has been suggested to be one reason for the enhanced thermal stability of *T. lanuginosus* and *P. varioti* xylanases [11,14]. A greater proportion of polar surface and a slightly extended C-terminus together with an extension of β-strand A5 are thought to increase the stability of *D. thermophilum* xylanase [15,19]. Despite all these studies, the structural basis for the thermostability of family 11 xylanases is not well understood.

We report here the three-dimensional structures of two new members of family 11 xylanases. The crystal structure of the catalytic domain from *Chaetomium thermophilum* xylanase Xyn11A (CTX) has been determined at 1.75 Å resolution and the catalytic domain from *Nonomuraea flexuosa* xylanase Xyn11A (NFX) at 2.1 Å resolution. CTX

Correspondence to N. Hakulinen, Department of Chemistry, University of Joensuu, PO Box111, FIN-80101 Joensuu, Finland. E-mail: Nina.Hakulinen@joensuu.fi

Abbreviations: AKX, *Aspergillus kawachii* xylanase; ANX, *Aspergillus niger* xylanase; BAX, *Bacillus agaradhaerens* xylanase; BCX, *Bacillus circulans* xylanase; CTX, *Chaetomium thermophilum* xylanase; DTX, *Dictyoglomus thermophilum* xylanase; GlcNAc, *N*-acetyl-D-glucosamine; NFX, *Nonomuraea flexuosa* xylanase; PVX, *Paecilomyces varioti* xylanase; THX, *Trichoderma harzianum* xylanase; TLX, *Thermomyces lanuginosus* xylanase; TRX I, *Trichoderma reesei* xylanase I; TRX II, *Trichoderma reesei* xylanase II.

Enzymes: xylanases (EC 3.2.1.8).

Note: The coordinates of the refined structures have been deposited with the Protein Data Bank, accession codes are 1H1A for *Chaetomium thermophilum* and 1M4W for *Nonomuraea flexuosa*.

(Received 1 November 2002, revised 17 January 2003, accepted 3 February 2003)

and NFX act optimally at 65–80 °C; NFX, in particular, is a remarkably stable enzyme, having a half-life of 273 min at 80 °C and even 28 min at 100 °C. In addition, NFX is active at pH 8. The crystal structures of CTX and NFX allowed us to make detailed comparison of 12 xylanases, five from thermophilic organisms. This gives a more reliable comparison of the enzyme structures in relation to their thermostability than earlier studies and helps us to understand the molecular basis of the thermostability of these industrially relevant enzymes.

Materials and methods

Protein purification

The catalytic domains of *C. thermophilum* and *N. flexuosa* expressed from *Trichoderma reesei* were purified from samples kindly provided by A. Mäntylä (ROAL, Rajamäki, Finland). GenBank accession codes are AJ508931 and AJ508952 for *C. thermophilum* xylanase and *N. flexuosa* xylanase, respectively. *C. thermophilum* xylanase was expressed in *T. reesei* as a full-length enzyme containing 235 amino acids and *N. flexuosa* as a construct coding mainly the catalytic domain (220 amino acids). However, it is likely that an extracellular protease has cleaved off the C-terminal tail of *C. thermophilum* xylanase (shortening was seen in SDS/PAGE) and possibly also several C-terminal residues outside the catalytic core of *N. flexuosa* xylanase. As determined by SDS/PAGE, the *C. thermophilum* xylanase was present as a $\approx$ 26 kDa protein and the *N. flexuosa* xylanase as $\approx$ 28 kDa. Both xylanases were purified by cation exchange chromatography (CM Sepharose Fast Flow; Amersham-Pharmacia Biotech, Uppsala, Sweden) and hydrophobic interaction chromatography (Phenyl Sepharose Fast Flow, Amersham-Pharmacia Biotech). The procedure was essentially the same as that described for *T. reesei* xylanase [20]. The *C. thermophilum* xylanase was further purified on a Q Sepharose High Performance column (Amersham-Pharmacia Biotech), equilibrated with 10 mM citrate buffer (pH 4). A linear gradient of 0–0.25 M NaCl in 10 mM citrate buffer (pH 4) was used to elute the enzyme.

Enzyme assay

The half-life of each xylanase was determined at different temperatures in 50 mM citrate/phosphate buffer, 0.01 mg·mL^{-1} bovine serum albumin, pH 6.0. After incubations at each temperature, the residual activity of xylanase was determined by measuring the amount of reducing sugars liberated from 1% birchwood xylan [21]. The half-lives were determined for enzymes produced in *T. reesei*.

Crystallization and data collection

The catalytic domains of *Chaetomium* xylanase (CTX) and *Nonomuraea* xylanase (NFX) were crystallized by a hanging-drop vapor-diffusion method at room temperature. CTX crystals were obtained in 8 mL droplets containing approximately 5 mg·mL^{-1} protein (A_{280} = 1 corresponds to the concentration 0.37 mg·mL^{-1} of protein) 0.7 M ammonium sulfate and 0.05 M Hepes at pH 7.2. Crystals of NFX were grown in 8 mL droplets containing 7 mg·mL^{-1} protein, 0.4 M ammonium sulfate and 0.05 M sodium acetate at pH 6.0. In both cases, the droplets were equilibrated against reservoir solution with a twofold higher concentration of ammonium sulfate and buffer. When sodium acetate buffer was used instead of Hepes, CTX also crystallized at pH 6–7, but these crystals diffracted only up to 3–4 Å. Similarly with Hepes buffer, NFX crystallized at pH 7–8, but the crystals were not suitable for X-ray analysis. High quality crystals of CTX (dimensions 0.5 × 0.2 × 0.2 mm) and NFX (0.3 × 0.2 × 0.2 mm) appeared in the drop after a few days and reached their final size in two weeks.

Data were collected on a Rigaku RU-200HB rotating anode X-ray generator operating at 50 kV and 100 mA equipped with an Osmic Confocal Optics and an RAXIS-IIC imaging plate detector. Initially, the data sets of CTX and NFX crystals were collected at room temperature at resolutions of 2.4 and 2.3 Å, respectively. Later, higher resolution data sets were collected at 120 K at resolutions of 1.75 Å and 2.1 Å, respectively. Crystals from both xylanases were soaked in cryoprotectant solution containing 30% glycerol. The diffraction images were processed with DENZO software and the data were scaled with SCALEPACK software [22]. The space groups were defined using XPREP program (SHELX software package). CTX crystals belonged to the orthorhombic space group $P2_12_12$ with unit cell parameters a = 108.24 Å, b = 57.15 Å, and c = 65.68 Å. The asymmetric unit contained two molecules. NFX crystals were hexagonal with unit cell parameters a,b = 37.03 Å, and c = 191.81 Å and they belonged to the space group $P6_1$. The asymmetric unit of NFX crystals contained only one molecule. The data collection statistics are presented in Table 1.

Structure solution and refinement

Both structures were determined using the molecular replacement method with the AMoRe program [23]. The search model was *Trichoderma reesei* xylanase II (TRX II, PDB code 1ENX). Sequence identities of CTX and NFX (digestion site determined with mass spectrometer) with TRX II are 63% and 51%, respectively. The molecular replacement solutions were initially calculated from the room temperature data sets and the models were further improved with the high-resolution data sets. Iterative cycles of refinement and manual fitting were carried out using programs CNS [24] and O [25]. To monitor the progress of the refinement, a total of 10% of the reflections were set aside for the R-free calculations. Refinements were carried out using the maximum-likelihood method with bulk-solvent corrections. The water molecules of the CTX model were positioned automatically with the wARP [26] but were also checked and finalized with the O. The water molecules of NFX were positioned with CNS and O. Refinement statistics of the final models are presented in Table 1. The CTX model contained four sulfates and two of them at special positions. As the refinement programs are not able to refine covalent bonds at special positions, only the sulfur atoms of these two sulfates were modeled. However, the electron density map showed clearly the

Table 1. Data collection and refinement statistics.

	CTX	NFX
Data collection		
Resolution range (Å) (overall/last shell)	99–1.75 (1.81–1.75)	99–2.1 (2.18–2.10)
No. of total observations	190677	48587
No. of unique reflections	40787	8541
I/I(σ)	27.1	17.9
Completeness of data (%) (overall/last shell)	97.3 (90.9)	97.6 (92.8)
R_{sym} (%) (overall/last shell)	7.7 (30.2)	10.7 (29.3)
Refinement		
Resolution range (Å)	99–1.75	99–2.1
No. of reflections F > 0 σ	39243	8135
R_{factor} (%)	17.9	14.6
R_{free} (%)	21.6	20.9
No. of non-hydrogen atoms	3637	1787
Protein	3015	1544
Water	603	183
Carbohydrate	–	50
Ligand (glycerol)	6	6
Ion	13	4
Average B (Å^2)	20.7	19.9
Main chain	16.5	17.1
Side chain	18.6	18.2
Water	35.8	31.6
Carbohydrate	–	42.3
Ligand (glycerol)	29.3	32.6
Ion	67.9	27.0
Rmsd from the ideal		
Bond lengths (Å)	0.005	0.006
Bond angles (°)	1.475	1.439

shape of the tetrahedral corresponding to the sulfate ion. In the final model, two conformations of residues A10, B10 and A123 of CTX were refined. There were also signs of other conformations of surface residues A32, A37, A38, A62, A70, A83, A110, B32, B37, B57, and B70 of CTX. The final model was evaluated with PROCHECK software [27].

Comparison of family 11 xylanases

The coordinates of 10 different family 11 xylanases were available in PDB. First, the Cα atoms of all xylanase models were roughly superimposed with the O program. To create the final multiple alignment, STAMP software [28] was applied. The secondary structures were assigned with the DSSP software [29].

The solvent accessible surface areas of xylanases were calculated with the NACCESS software [30] using a 1.4 Å probe. Van der Waals volumes were calculated using 1.4 Å probes and without probes with the VOIDOO program [31]. Numbers of hydrogen bonds were calculated for all xylanases using HBPLUS routine [32] with the default parameters for distances and angles. A salt bridge was assigned, when the distance between the two atoms of opposite charge was less than 4 Å [33]. In all calculations, water molecules and hetero-atoms were excluded from the coordinate files and the chains were split.

Mass spectrometry

Mass spectra of CTX and NFX were measured by a Bruker BioAPEX II 47e FTICR mass spectrometer (Bruker Daltonics) using positive mode electrospray ionization (ESI). This instrument is equipped with a passively shielded 4.7-T superconducting magnet, cylindrical infinity ICR cell and external electrospray ion source (Analytica of Branford). Aliquots of CTX and NFX were diluted with a methanol/water (1 : 1, v/v) solvent, followed by glacial acetic acid (1%) to obtain denaturing solution conditions for efficient protonation in ESI. The final concentration for both proteins was approximately 0.5 mg·mL^{-1}. Samples were infused into the ESI-source by a syringe infusion pump (Cole-Parmer) at a rate of 50 μL·h^{-1}. Ionization voltage was −3.7 kV and ions were accumulated for 2 s in an RF-only hexapole ion guide before they were transferred into the ICR cell for excitation and detection. The drying gas in a spraying process was pure nitrogen gas. All data were acquired and processed with a Bruker XMASS 5.0.1. software. The mass spectra were calibrated against an acetonitrile-based ES Tuning Mix (Hewlett Packard) by peptide peaks in the m/z range of 200–3000. Molecular masses of observed proteins were calculated as average values over the charge state distributions using the ESIMASS macro program. Relative abundances of glycosylated and nonglycosylated protein species were calculated

using the absolute intensities of the peaks appearing in the ESI-spectra.

Results and discussion

Overall structure of *C. thermophilum* xylanase

The overall structure of xylanase from *C. thermophilum* (CTX) was dominated by one α-helix and two strongly twisted β-sheets, which were packed against each other. This is the protein fold of family 11 xylanases. According to Törrönen *et al.* [8], the shape of the molecule resembles a right hand: two β-sheets and the α-helix form fingers and a palm, a long loop between the B7 and B8 strands forms a thumb, and a loop between the B6 and B9 strands forms a cord (Fig. 1A).

The final model of CTX contained residues 1–191 for both molecules in the asymmetric unit (labeled A and B). The first residue, glutamine, was deaminated and cyclized to pyrrolidone carboxylic acid. When the ESI mass spectrum of CTX was measured, the unique molecular masses, 21 479 Da and 21 682 Da, were obtained. Assuming that CTX contains 196 residues, the calculated molecular mass would be 21 478 Da, which agrees well with the lower mass obtained. The difference between the two obtained masses was 203 Da corresponding to one *N*-acetyl-glucosamine (GlcNAc). There is one potential *N*-glycosylation site (Asn62) in the sequence of CTX, but there was no clear sign of glycosylation in the electron density map. It is possible that only the protein molecules without GlcNAcs had been crystallized or that the GlcNAc is disordered. According to the mass spectrum, approximately 20% of the material did not contain GlcNAc or alternatively, the GlcNAc had been lost.

In the crystal structure, a glycerol molecule was located in the active site of molecule A, but was not observed in molecule B. The cryoprotectant soaking solution was most likely the source of glycerol, which was packed against Trp19 by stacking interactions and was hydrogen-bonded to carboxyl group of Pro127. In addition, Arg123 had two conformations in molecule A and in one of the conformations the guadinine group of the Arg was located towards the hydroxyl group of the glycerol. The rms deviation between the A and B molecules of CTX was 0.8 Å.

The crystal structure showed four sulfate ions and a calcium ion in the asymmetric unit. The calcium ion was located between molecules A and B exactly on the noncrystallographic axis. The calcium ion interacted with side chains Oγ of Thr10 from molecule A and B, both of which clearly had two conformations in the electron density map. Two of the sulfate ions were located exactly on the crystallographic axes. In addition to these two sulfate ions, which are attached to Arg residues A27 and B27, there are two other sulfates, which are attached to Arg residues A68 and B68. Due to the crystal packing, the enzyme resembles a tetrameric assembly (Fig. 1B). Four sulfate ions link molecule A to symmetry molecule D and correspondingly molecule B to symmetry molecule C. However, according to the dynamic light scattering measurements, the protein was a monomer. Therefore, the sulfate ions from the crystallization solution might have been involved in this 'tetramerization' process. It is possible that tetramers were assembled first and their stacking then led to crystal formation in the high salt concentration.

Overall structure of *N. flexuosa* xylanase

The protein fold of the xylanase from *N. flexuosa* (NFX) was the same as that of CTX and other family 11 xylanases (Fig. 2A). The crystal structure of NFX contained 197 residues with a sequence GNPGNP at the C-terminus. The sequence GNPGNP seems to be a part of the C-terminal tail of the full-length NFX with total 301 residues. The amino acids of the C-terminal tail were sticking out from the model, but if the C-terminus of NFX is excluded, the enzyme is slightly more compact than CTX, probably due to short deletions. The electron density map showed that there was Ala at position 73 instead of Gly, which had been

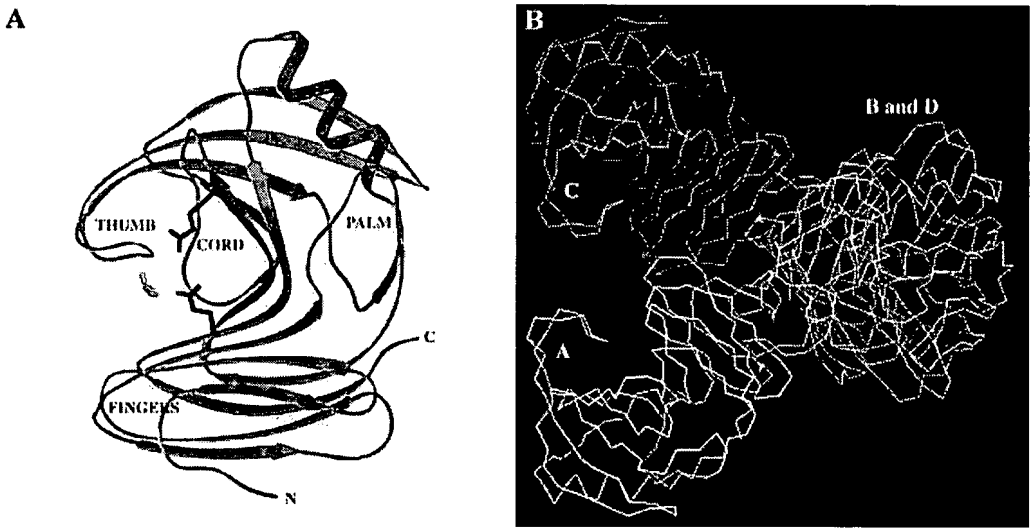


Fig. 1. CTX. (A) The overall structure of CTX. Glycerol and catalytic glutamates are shown in the active site. (B) A tetrameric assembly with sulfate ions. Molecules A and B are shown in white and symmetry molecules C and D in blue.

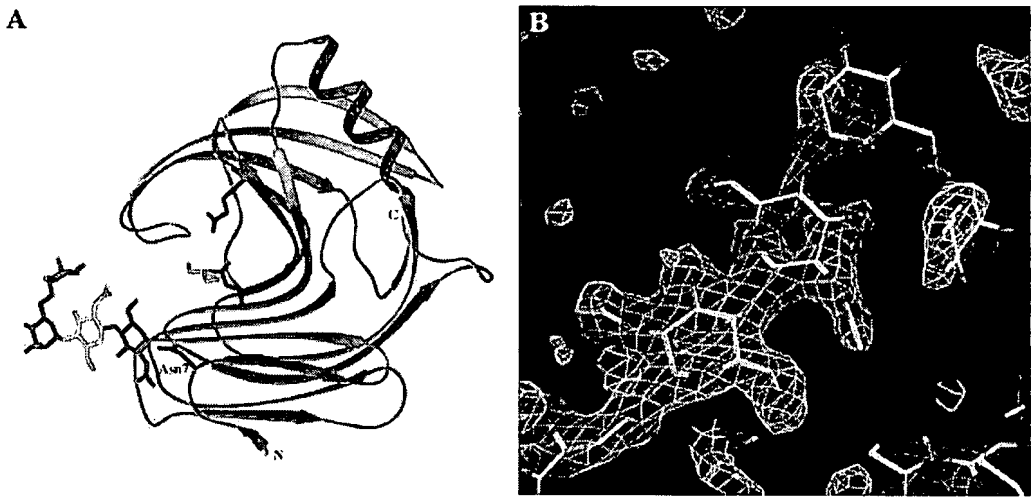


Fig. 2. NFX. (A) The overall structure of NFX with a glycerol molecule in the active site. Carbohydrates attached to Asn7 are shown in gray sticks. (B) The representative $2F_o–F_c$ electron density map from the final model of NFX. The figure shows the density of carbohydrates, contoured at a level of 1.5 σ.

determined earlier by sequencing. This might be a sequencing error or mutation in the *T. reesei* strain.

The crystal structure of NFX revealed that the enzyme has a single *N*-glycosylation site at Asn7, where two *N*-acetyl-glucosamines and two mannoses were attached (Fig. 2B). Carbohydrates are orientated in the same way as the backbone of β-strand B1 and therefore they are almost like an extension of the β-strand. *N*-glycans are known to have a stabilizing effect and they may prevent the aggregation of unfolded protein molecules [34]. However, NFX contains *N*-glycans only when the enzyme is expressed in *T. reesei*. In the ESI mass spectrum, we were able to see six peaks in every charge state, corresponding to heterogeneously glycosylated molecules. From the peaks of the most abundant charge state distribution corresponding to a mass of 23491 Da, we concluded that most of the material contained 206 residues, two GlcNAcs and three mannoses. The difference of 22 Da between the calculated (22577 Da) and the measured (22599 Da) was probably due to the formation of sodium adduct.

On the protein surface, an acetate ion was located 3.2 Å away from Ser187 Oγ and 2.8 Å from Ser36 Oγ. The glycerol molecule was again found in the active site. It was slightly differently located in the active site of NFX than it was in the active site of CTX. The glycerol was packed against Trp20 (corresponding to Trp19 in CTX), but it was slightly deeper in the active site. In NFX, Tyr170 and Tyr78 interacted with hydroxyl groups of glycerol. When this complex structure is compared with the complex structure of *T. reesei* xylanase with epoxyalkyl xylosides [35], the glycerol appears to mimic the binding of the xylose ring in the active site. The binding of glycerol to a single site may suggest that this site is the strongest binding subsite for the xylose subunits of xylan.

Structural comparison of family 11 xylanases

The *C. thermophilum* (CTX) and *N. flexuosa* (NFX) xylanases are much more thermostable than the mesophilic *T. reesei* xylanase II (TRX II). While TRX II was rapidly inactivated at 55–60 °C, CTX was stable up to 60–65 °C and NFX was stable at 80 °C and it had some stability even at 90–100 °C (Table 2). However, the reasons for the significantly higher thermostability of NFX and CTX are not readily evident at the structural level, as they both resemble TRX II xylanase very closely. Because a number of solved three-dimensional structures of family 11 xylanases are now available for both mesophiles and thermophiles, we made a detailed comparison of the structures of these enzymes. Twelve structures used in the comparison are summarized in Table 3 and the sequence alignment is shown in Fig. 3.

According to the sequence homology of family 11 xylanases, which have a solved crystal structure, the enzymes can be divided into four groups (Fig. 4). The first group is formed from acidophilic xylanases ANX, AKX, and TRX I. The second group contains alkalophilic BAX and highly thermophilic DTX (sequence identity 58%). The third group is formed from thermophilic NFX and mesophilic BCX (sequence identity 59%). The fourth group contains mesophilic THX and TRX II together with thermophilic PVX, TLX, and CTX. BAX, DTX, BCX, and NFX are all from bacterial sources, whereas the others are fungal enzymes.

Table 2. Half-lives of TRX II, CTX, and NFX.

Temperature (°C)	TRX II (min)	CTX (min)	NFX (min)
50	1480	–	–
55	20	–	–
60	2	1500	–
65		58	–
70		15	–
75		7	1500
80		4	273
85			148
90			88
95			39
100			28

Table 3. Summary of the crystal structures used in comparison.

Organism	Code	PDB code	Temperature preference	pH preference	Resolution (Å)	Measurement T (K)	Ligand	Reference
N. flexuosa	NFX	1M4W	Thermophile		2.1	120	Glycerol	This paper
C. thermophilum	CTX	1H1A	Thermophile		1.75	120	Glycerol in molecule A	This paper
D. thermophilum	DTX	1F5J	Thermophile		1.8	110		[15]
T. lanuginosus	TLX	1YNA	Thermophile		1.55	295		[11]
P. varioti	PVX	1PVX	Thermophile		1.6	295		[14]
T. reesei	TRX II	1ENX			1.5	295		[8]
B. circulans	BCX	1XNB			1.5	295		[5]
T. harzianum	THX	1XND			1.8	295		[5]
T. reesei	TRX I	1XYN		Acidophile	2.0	295		[9]
A. kawachii	AKX	1BK1		Acidophile	2.4	295		[12]
A. niger	ANX	1UKR		Acidophile	2.0	295		[10]
B. agaradhaerens	BAX	1QH7		Alkalophile	1.8	100	β-D-xylopyranoside	[13]

When three-dimensional structures of family 11 xylanases are superimposed, their rmsd (root-mean-square deviation) values correlate well with the sequence similarities. The sequence identities of family 11 xylanases, including all molecules in the asymmetric unit, are shown in the function of rmsd values in Fig. 5. The sequence identity range for different family 11 xylanases was 31–97% and rmsd range was 0.2–1.4 Å. The natural structural differences can be seen in the upper part of the figure (sequence identity 100%). The high rmsd value 0.8 Å, which exists between molecules A and B of CTX, is partly due to movements induced by glycerol binding. Therefore, we note that some of the structural differences among family 11 xylanases are caused by ligand binding. Both NFX and molecule A of CTX contain the glycerol while BAX contains the β-D-xylopyranoside in the active site. For the structural comparisons, other family 11 xylanases were chosen without ligands.

The superimposition of three-dimensional structures confirmed the subgroups of xylanases based on sequence similarities. For example, the lowest rmsd value of thermophilic NFX is with mesophilic BCX (0.78 Å), both belonging to group 3. NFX had a low rmsd (0.82 Å) with molecule A of thermophilic CTX, showing that groups 3 and 4 are closely related. Molecule A of CTX has the lowest rmsd with mesophilic THX (0.71 Å) and molecule B of CTX with mesophilic TRX II (0.75 Å), all belonging to group 4. In group 1, alkalophilic BAX has the lowest rmsd with thermophilic DTX (0.79 Å between molecule A of BAX and molecule A of DTX).

As the crystal structures of mesophilic and thermophilic xylanases are very similar, it is likely that an array of minor modifications forms the structural basis for enhanced stability in thermophilic xylanases. Therefore, several factors, which are thought to be responsible for thermostability, were compared between thermophilic and mesophilic family 11 xylanases. The alkalophilic BAX was not included in the same group as other mesophilic xylanases, because its functional properties seem to be different. *Bacillus agaradhaerens* grows optimally at unusually high pH (over 10). On the other hand, acidophilic TRX I, AKX and ANX would be considered as a separate group of acidic xylanases, but in our comparisons they were included in the mesophiles as a large number of mesophilic xylanases are slightly acidic in their activity profiles. The C-terminal tail (GNPGNP) of NFX was excluded.

Sequence properties

Frequencies of all 20 amino acids were computed for thermophilic and mesophilic family 11 xylanases (Table 4). It is obvious that the comparison of amino acid contents suffered from the low number of sequences and thus, statistical methods were not used to analyze the data. However, this comparison may still reveal some important trends and some of the trends in the amino acid frequencies could be related to the thermostability of xylanases.

There was found an increased occurrence of arginines in the thermophilic xylanases. Large-scale sequence comparisons have shown that thermophilic proteins contain more arginines on the protein surface than mesophilic proteins [36–38]. The effect of the large-scale increase in the number of arginines was tested experimentally in *T. reesei* xylanase II [39]. These results showed that the introduction of five arginines into the Ser/Thr surface increased considerably the thermotolerance in the presence of the substrate.

Another trend is that in thermophilic xylanases the frequency of Ser decreases and correspondingly the frequency of Thr increases (Table 4). Ser → Thr mutation was one of the stabilizing mutations found by the early study of Argos *et al.* [40]. For thermophilic proteins, the decrease in the frequency of Ser but not the increase of Thr was observed by Kumar *et al.* [38]. These authors found that in thermophilic proteins, Arg and Tyr are more frequent, while Cys and Ser are less frequent. One possible explanation for this finding in xylanases is that the increase in the Thr : Ser ratio in β-strands (Table 5) improves the β-forming propensities. Over half of the residues in the family 11 xylanases are located in the β-strands.

In thermophilic xylanases, the frequency of asparagines is slightly lower (Tables 4 and 5). Asn has a low β-forming propensity, and thus might be avoided in the β-strands of thermophilic xylanases. The highly thermostable xylanase DTX showed a decreased frequency of Gly (Table 4)

```
                A1   B1       B2         A2          A3           B3
TRXI                  asiNYDQNyqt ggQVSYSp sNTGFSVnw  ntQDDFVVGVGWt
AKX                   agiNYVQNyngnLGDFTyd esAGTFSMywedgvSSDFVVGLGWt
ANX                    giNYVQNyngnLGDFTyd esAGTFSMywedgvSSDFVVGLGWt
BCX                   astDYWQNwtdgggIVNAVngsGGNYSVnw  snTGNFVVGKGWt
DTX          altsnasgtfdgYYYELwkd tgNTTMTvytQGRFSCqw  snINNALFRTGKk
BAX         qivtdnsignhdgYDYEFwkdsggSGTMIlnhGGTFSAqw  nnVNNILFRKGKk
NFX        dttitqnqtgydngYFYSFwtdapgTVSMTlhsGGSYSTsw  rnTGNFVAGKGWs
PVX          gttpnsegwhdgYYYSWwsdgggDSTYTnnsGGTYEItw  gnGGNLVGGKGWn
TLX          qttpnsegwhdgYYYSWwsdggaQATYTnleGGTYEIsw  gdGGNLVGGKGWn
THX         qqtigpgtgysngYYYSYwndghaGVTYTnggGGSFTVnw  snSGNFVAGKGWq
TRXII        qtiqpgtgynngYFYSYwndghgGVTYTngpGGQFSVnw  snSGNFVGGKGWq
CTX         qtltssatgthngYYYSFwtdgqgNIRFNlesGGQYSVtw  sgNGNWVGGKGWn

                          A5              B5          B6       CORD
TRXI    t         gsSAPINF gGSFSvNSGTGLLSVYGWSTNPLVEYYIMEDNHNYP  a
AKX     t         gsSNAITY sAEYSaSGSSSYLAVYGWVNYPQAEYYIVEDYGDYNpcs
ANX     t         gsSNAITY sAEYSaSGSASYLAVYGWVNYPQAEYYIVEDYGDYNpcs
BCX     t        gspFRTINYnaGVWA PNGNGYLTLYGWTRSPLIEYYVVDSWGTYRp
DTX     ynqn  wqslgTIRITY sATYN PNGNSYLCIYGWSTNPLVEFYIVESWGNWRp p
BAX     fnetqthqqvgNMSINY gANFQ PNGNAYLCVYGWTVDPLVEYYIVDSWGNWRp p
NFX     t         ggRRTVTY nASFN PSGNAYLTLYGWTRNPLVEYYIVESWGTYRp
PVX     p        glnARAIHF tGVYQ PNGTSYLSVYGWTRNPLVEYYIVENFGSSNpss
TLX     p        glnARAIHF eGVYQ PNGNSYLAVYGWTRNPLVEYYIVENFGTYDpss
THX     p        gtkNKVINF sGSYN PNGNSYLSIYGWSRNPLIEYYIVENFGTYNpst
TRXII   p        gtkNKVINF sGSYN PNGNSYLSVYGWSRNPLIEYYIVENFGTYNpst
CTX     p        gtdNRVINY tADYR PNGNSYLAVYGWTRNPLIEYYVVESFGTYDpst

              B9          B8       THUMB      B7              A6
TRXI    QGTVKGTVTSDGATYTIWENTRVNEPSI qGTATFNQYISVRNspr    tSGTVTVQ
AKX     SATSLGTVYSDGSTYQVCTDTRTNEPSI tGTSTFTQYFSVREstr    tSGTVTVA
ANX     SATSLGTVYSDGSTYQVCTDTRTNEPSI tGTSTFTQYFSVREstr    tSGTVTVA
BCX     TGTYKGTVKSDGGTYDIYTTTRYNAPSIdgDRTTFTQYWSVRQskrptgsNATITFT
DTX     GATSLGQVTIDGGTYDIYRTTRVNQPSI vGTATFDQYWSVRTskr    tSGTVTVT
BAX     GATPKGTITVDGGTYDIYETLRVNQPSI kGIATFKQYWSVRRskr    tSGTISVS
NFX     TGTYKGTVTTDGGTYDIYETWRYNAPSI eGTRTFQQFWSVRQqkr    tSGTITIG
PVX     GSTDLGTVSCDGSTYTLGQSTRYNAPSI dGTQTFNQYWSVRQdkr    sSGTVQTG
TLX     GATDLGTVECDGSIYRLGKTTRVNAPSI dGTQTFDQYWSVRQdkr    tSGTVQTG
THX     GATKLGEVTSDGSVYDIYRTQRVNQPSI iGTATFYQYWSVRRnhr    sSGSVNTA
TRXII   GATKLGEVTSDGSVYDIYRTQRVNQPSI iGTATFYQYWSVRRnhr    sSGSVNTA
CTX     GATRMGSVTTDGGTYNIYRTQRVNAPSI eGTKTFYQYWSVRTskr    tGGTVTMA

         HELIX              B4          A4
TRXI    NHFNAWASLGL hLGQMNYQVVAVEGWGGSGSAsqsvsn
AKX     NHFNFWAQHGF gNSDFNYQVMAVEAWSGAGSAsvtis
ANX     NHFNFWAHHGF gNSDFNYQVVAVEAWSGAGSAsvtis
BCX     NHVNAWKSHGMnlGSNWAYQVMATEGYQSSGSSnvtvw
DTX     DHFRAWANRGL nLGTIDQITLCVEGYQSSGSAnitqntfsqss
BAX     NHFRAWENLGM nMGKMYEVALTVEGYQSSGSAnvysntlringnpls
NFX     NHFDAWARAGM nLGSHDYQIMATEGYQSSGSStvsiseggnpgnp
PVX     CHFDAWASAGLnvTGDHYYQIVATEGYFSSGYAritvadvg
TLX     CHFDAWARAGLnvNGDHYYQIVATEGYFSSGYAritvadvg
THX     NHFNAWASHGL tLGTMDYQIVAVEGYFSSGSAsitvs
TRXII   NHFNAWAQQGL tLGTMDYQIVAVEGYFSSGSAsitvs
CTX     NHFNAWRQAGL qLGSHDYQIVATEGYYSSGSAtvnvg
```

Fig. 3. Sequence alignment of family 11 xylanases. Structurally very similar residues are in capital letters. The coloring (red, α-helix; green, 3_{10}-type helix; blue, β-strand) depicts the secondary structure elements.

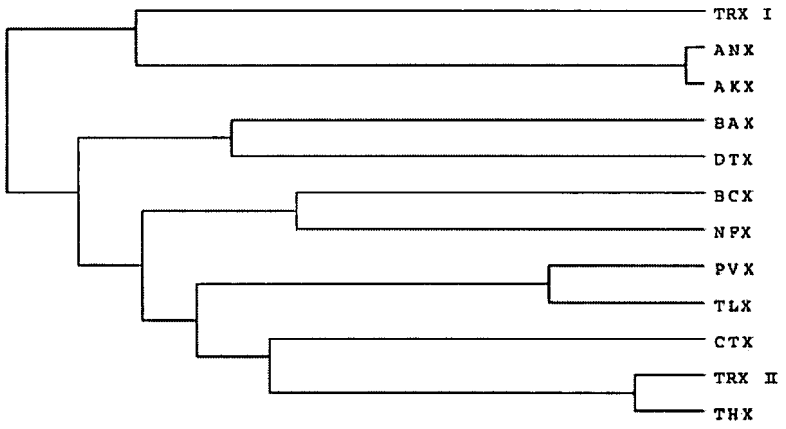


Fig. 4. Phylogenetic tree of family 11 xylanases. Lengths of branches indicate the evolutionary distances.

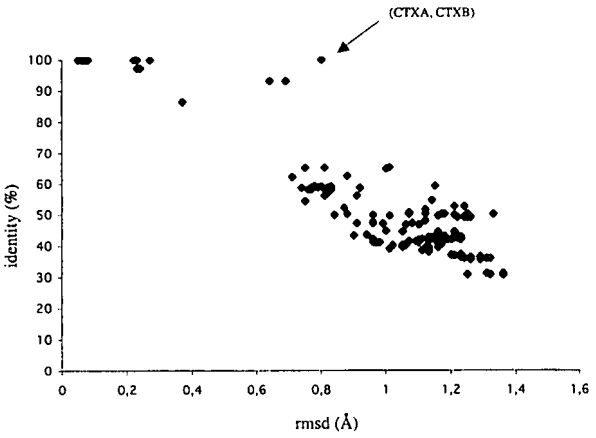


Fig. 5. The plot of sequence identity as a function of rmsd value for family 11 xylanases.

compared to both mesophilic and other thermophilic xylanases. Avoidance of Gly probably increases the rigidity of the loop regions. However, there is no general trend toward decreased frequency of Gly among thermophilic family 11 xylanases. Pro does not seem to play any general role in the thermostabilization of these enzymes.

Thermophilic xylanases have substantially less Val (Tables 4 and 5). Although Val has a good β-forming propensity, still its frequency is lower in the β-strands of the thermophilic xylanases (Table 5), indicating the increase in the β-forming propensity is not of primary importance in xylanases if some other property is more critical for thermostability. In addition, thermophilic xylanases contain more amino acid residues in the solved crystal structures than mesophilic xylanases (Table 4). The higher frequency of charged residues is involved in increasing the number of polar interactions.

Secondary structures

Facchiano *et al.* [41] observed that 69% of the α-helices of thermophilic proteins are more stable than their mesophilic counterparts. The stabilizing factor was the intrinsic helical propensity of amino acids. Lack of β-branched residues (Val, Thr, Ile) correlated significantly with thermostability. In the case of xylanases, there is only one α-helix in the structure. The α-helix of thermophilic xylanases showed a higher frequency of Asp and Arg (Table 5). In NFX, the additional Arg160 is located on the protein surface, and Asp156 makes a double salt bridge with Arg58 (Oγ1 and Oγ2 atoms of aspartic acid and Nη1 and Nη2 atoms of arginine). In CTX, Arg161 makes a salt bridge with Asp57. Hence, the α-helix region of these two enzymes is likely to be stabilized by additional interactions with the loop before the β-strand A5. DTX, NFX and TLX have both Asp and Arg in the α-helix and the residues are located at positions (i, i + 3) or (i, i + 4), which is believed to be stabilizing [42]. In addition, CTX has Met and Phe side chains at positions (i, i + 4), also thought to be stabilizing [43]. In NFX, TLX and CTX, the positively charged Arg is located at the C-terminal end of the α-helix, suggesting that it stabilizes the helix dipole [44].

The total number of positions with a β-strand structure was higher in thermophilic xylanases (Table 5). The thermophilic xylanases had, on average, 123 residues (range 121–128) in the β-strands and the corresponding number in the mesophilic xylanases was 114 residues (range 106–118). This result indicates that longer β-strand rigidify the protein and, thus, make it more thermostable. Alkalophilic BAX had as many as 131 residues in its β-strands, which indicates that the overall stability of the β-strands may be important for the alkalitolerance of family 11 xylanases. All thermophiles and BAX have an additional β-strand B1 at the N-terminus, which could have a stabilizing effect. However, mesophilic TRX II and THX also have this additional β-strand. The highly thermostable DTX has a clearly longer β-strand B3 and C-terminal β-strand A4, which most likely stabilize the structure. The C-terminal β-strand A4 gives additional hydrogen bonding with β-strand A5 and the extension of β-strand B3 interacts with a β-strand B4. BAX also has a longer C-terminal β-strand A4 and a short additional β-strand after that.

When the three-dimensional structures of all xylanases are superimposed, a striking feature, in addition to the lengths of the terminals, is that thermostable DTX and alkalophilic BAX have a long insertion between β-strands B3 and A5. According to McCarthy *et al.* [15], the loop between B3 and A5 combined with extended C-terminus of DTX gives additional hydrogen bonding and hydrophobic

Table 4. Total amino acid composition. The largest differences between thermophiles and mesophiles are in bold.

	NFX	CTX	DTX	TLX	PVX	thermo	TRX II	BCX	THX	TRX I	AKX	ANX	meso	BAX
% Ala	4.2	5.2	4.5	6.7	4.6	5.1	3.7	4.8	4.7	5.1	7.6	8.2	5.7	3.9
% Val	4.2	6.8	5.0	6.7	6.2	**5.8**	7.4	7.5	6.8	10.1	8.2	8.7	**8.1**	6.8
% Leu	2.6	3.1	5.0	4.1	3.6	3.7	2.6	2.1	2.6	3.4	2.2	2.2	2.5	4.3
% Ile	4.2	3.1	5.5	3.6	3.1	3.9	4.7	3.2	5.3	3.4	2.7	2.7	3.7	5.8
% Pro	2.6	2.6	2.5	3.1	3.1	2.8	3.7	3.2	3.2	3.4	1.6	1.6	2.8	3.4
% Met	1.6	1.0	0.5	0.0	0.0	0.6	0.5	1.1	0.5	1.1	1.1	0.5	0.8	2.4
% Phe	3.7	2.6	3.5	2.6	2.6	3.0	4.2	2.1	3.7	3.4	4.9	4.9	3.9	3.4
% Trp	4.2	3.7	4.0	4.1	4.1	4.0	3.2	5.9	3.2	3.4	2.7	2.7	3.5	3.4
% Gly	13.6	14.1	9.5	14.9	16.0	13.6	14.2	13.4	14.2	12.4	10.3	10.3	12.5	11.6
% Ser	10.5	8.9	10.1	6.7	11.3	**9.5**	11.6	9.6	12.6	12.9	15.8	15.2	**13.0**	8.2
% Thr	15.2	12.6	14.1	9.3	10.8	**12.4**	8.4	13.4	8.9	10.1	10.9	10.9	**10.4**	7.7
% Cys	0.0	0.0	1.5	1.0	1.0	0.7	0.0	0.0	0.0	0.0	1.1	1.1	0.4	0.5
% Tyr	8.4	9.4	7.0	8.8	8.8	8.5	8.9	8.0	9.5	5.6	9.2	9.2	8.4	6.3
% Asn	5.8	8.4	8.5	6.2	7.2	**7.2**	10.5	9.6	10.0	10.1	6.5	6.5	**8.9**	10.6
% Gln	4.2	4.2	5.5	4.1	3.6	4.3	5.3	2.7	3.2	6.2	3.3	2.7	3.9	4.3
% Asp	3.7	3.1	3.5	6.2	5.2	4.3	2.1	3.7	2.1	2.8	4.9	4.9	3.4	3.9
% Glu	3.1	2.6	2.0	4.1	2.6	2.9	2.1	1.1	2.1	2.8	4.3	4.3	2.8	3.4
% Lys	1.6	1.6	2.0	1.5	1.0	1.5	2.1	2.7	2.1	0.6	0.0	0.0	1.2	4.3
% Arg	5.2	5.2	5.0	4.1	3.1	**4.5**	3.2	3.7	3.2	1.7	1.6	1.6	**2.5**	3.9
% His	1.6	1.6	0.5	2.1	2.1	1.6	1.6	1.1	2.1	1.7	1.1	1.6	1.5	1.9
Total number	191	191	199	194	194	**194**	190	187	190	178	184	184	**186**	207
% Non-polar	27.7	27.5	30.6	30.9	27.3	28.8	30	31	30	33.1	31	31.5	31.1	33.3
% Polar	58.2	58.7	56.3	51	58.8	56.6	58.9	56.7	58.4	57.3	57.1	56	57.4	49.3
% Charged	14.1	13.8	13.1	18.1	13.9	**14.6**	11.1	12.3	11.6	9.6	11.9	12.5	**11.5**	17.4

Table 5. Amino acid composition in α-helices and β-strands. The largest differences between thermophiles and mesophiles are in bold.

	α-Helices		β-Strands	
	Thermophiles	Mesophiles	Thermophiles	Mesophiles
Ala	2.0	2.0	5.4	5.3
Val	0.2	0.7	**9.8**	**14.2**
Leu	–	–	3.8	2.5
Ile	0.2	–	6.2	4.8
Pro	–	–	1.4	1.3
Met	0.2	–	0.8	1.3
Phe	1	1.3	4.4	5.0
Trp	1	1	6.4	5.0
Gly	0.6	–	11.0	9.8
Ser	0.2	0.5	12.0	13.0
Thr	0.8	0.5	**16.8**	**12.8**
Cys	0.4	–	1.0	0.3
Tyr	–	–	15.6	14.3
Asn	0.6	2.0	**6.0**	**8.0**
Gln	0.2	0.3	5.4	4.5
Asp	**0.8**	–	3.0	3.0
Glu	–	–	4.8	4.3
Lys	–	0.2	2.0	1.3
Arg	**0.8**	–	5.0	3.0
His	1	1	1.8	0.2
Total	10	9.5	**122.6**	**114.2**

packing. They suggest that these factors may account for the enhanced thermal stability. In fact, the loop of DTX includes regular secondary structures: the extension of β-strand B3 and 3_{10}-type helix. Structurally similar BAX has a short helix instead of 3_{10}-type helix, but there is no extension of β-strand B3.

Furthermore, the structures of cord, thumb and loop regions vary among family 11 xylanases. It has been shown that these areas are flexible both in crystals and in the molecular dynamics simulations [45]. Some of the differences are evident in the ligand binding [35]. Loops are typically the regions with the largest temperature factors, indicating that they might unfold first during thermal denaturation [46]. However, the overall temperature factors of mesophilic and thermophilic xylanases were not comparable because the data sets of different xylanases have been collected at 120 K or at room temperature. Temperature factors are dependent, in addition to the resolution and the programs used in the refinements, on the temperature of crystal during data collection.

Several xylanases have short insertions or deletions in the loops (Fig. 3). Mesophilic BCX has a short insertion between β-strands B7 and A6 and mesophilic ANX and AKX have short insertions between β-strands A3 and B3, but there is no clear trend that shortened loops would be associated with thermostability. The loop between β-strands B2 and A2 has an interesting feature that could play a role in thermostability. The thermophilic NFX has Pro in this loop, which increases the rigidity and this might have a stabilizing effect. The other highly thermostable xylanase, DTX, has a deletion in this loop.

Disulfide bridges

Thermophilic PVX and TLX have a disulfide bridge that connects the C-terminus of the β-strand B9 with the N-terminus of the α-helix. According to the experimental

data, introducing disulfide bridges via site-directed mutagenesis has increased the thermostability in *T. reesei* and *B. circulans* xylanases. Disulfide bridges at the N-terminus or in the α-helix region improve the thermostability by 10–15 °C [20,46–48]. However, the disulfide bridge alone cannot be crucial for the enhanced thermal stability of xylanases due to the fact that highly thermostable DTX, NFX and CTX do not contain disulfide bonds. In DTX, two of the cysteines are close enough to form a disulfide bridge between β-strands B5 and B4 in the catalytic area, but are not reported to do that according to the electron density map [15]. In addition, the mesophilic AKX and ANX have a disulfide bond between the cord and the β-strand B8, indicating that stabilization of the α-helix region as well as other weak areas like N-terminus by various strategies is more important for the thermostability than the disulfide bridge alone.

Salt bridges and hydrogen bonds

There are increasing data that indicate a role for hydrogen bonds and salt bridges in protein stabilization [37,38]. In family 11 xylanases, the number of salt bridges varies between 2 and 12 (Table 6). There is one completely conserved salt bridge between the C-terminal Glu (or Asp) of β-strand B6 (BAX and BCX have Asp) and Arg of the loop between β-strands B7 and A6. Thermophiles tend to have more salt bridges, but on the other hand mesophilic TRX II has as many as eight salt bridges. Alkalophilic BAX has the largest number of salt bridges, while acidophilic xylanases have the lowest numbers. Apparently, there could be a correlation between alkalitolerance and salt bridges.

Thermophilic xylanases have, on average, slightly more hydrogen bonds than mesophilic xylanases, except the total number of hydrogen bonds in thermophilic CTX is lower than that of mesophilic TRX II (Table 6). Thermophiles, especially NFX, have a large number of side chain–side chain interactions.

Packing

It has been proposed that thermophilic proteins have a tighter internal packing with smaller and less numerous cavities than mesophilic proteins [49,50]. To study packing, we calculated the protein density and the void volume values for all family 11 xylanases (Table 7). Because thermophilic xylanases have more atoms, the void volumes were divided by the total number of atoms to normalize them. Both the protein density and void volume values for thermophilic and mesophilic xylanases were similar. Only highly thermophilic DTX and alkalophilic BAX have slightly higher protein density and lower void volumes indicating better packing.

In the comparison of PDB structures, Karshikoff and Ladenstein [51] have observed that proteins from thermophilic and mesophilic organisms essentially do not differ in packing. They suggest that neither the reduction in packing density nor the reduction of the packing defects can be considered as a common mechanism for increasing thermal stability. On the other hand, Chen *et al.* [52] observed in the mutagenesis study that the stabilizing mutations in *Staphylococcal* nuclease resulted in improved packing, with the volume of the mutant protein's hydrophobic cores decreasing as protein stability increased. Apparently, a few protein families or some members in them may use better packing to improve the thermostability. Our study indicated that highly thermostable DTX may benefit from the better packing. Adaptation to alkaline pH might also benefit from better packing.

Hydrophobicity and surface characteristics

Because protein cores are typically hydrophobic, increased packing efficiency is often correlated with increased hydrophobicity. Tighter packing can be achieved through the formation of hydrophobic clusters and enhanced van der Waals interactions. Increased hydrophobicity is usually involved in decreased accessible surface areas and a higher percentage of buried atoms [53]. According to our calculations (data not shown), thermophilic xylanases have slightly more apolar interactions on average than mesophilic xylanases, but if the number of interactions are divided with the number of residues, the trend is not as clear anymore.

As thermophilic xylanases contains more amino acid residues than mesophilic xylanases, they also have larger accessible surface areas (Table 7). So far, all family 11 xylanases are reported to be monomers, therefore the solvent accessible areas are not buried by oligomerization. When accessible surface area is counted per atom, it appears that DTX and BAX may benefit from increased hydrophobicity as well as better packing. In addition, these two xylanases have on average longer side-chains (atoms per residue) than the other family 11 xylanases studied (Table 7).

One type of hydrophobic interaction is the closely packed aromatic ring–ring interaction, which has been calculated to have nonbonded potential energies between 1 and 2 $kcal \cdot mol^{-1}$ [54]. Additional aromatic–aromatic interactions are believed to contribute to the increased stability [55]. *Bacillus* D3 xylanases, which belong to family 11 (no PDB coordinates available), have eight additional surface aromatic residues which are believed to form 'sticky patches' on the protein surface that may lead to protein aggregation [16]. In addition, introduction of a single tyrosine into the N-terminal region has been reported to improve the thermostability and thermophilicity of *Streptomyces* xylanase considerably [56]. However, the studied family 11 xylanases did not show any general trend toward increased proportion of aromatic residues (Table 4).

It is thought that increased fractional polar surface, which results in added hydrogen bonding to water, contributes to the greater stability [37]. Table 7 shows the solvent accessible areas and the fractions of polar and nonpolar surface areas. Thermophilic xylanases have somewhat larger fractional polar surfaces, especially CTX and DTX. This indicates that polar interactions on the protein surface are important for the stabilization of family 11 xylanases.

Conclusions

It appears from the analysis of three-dimensional structures and sequence properties of family 11 xylanases that there

Table 6. Polar interactions. Number of interactions/number of residues are given in parenthesis. The largest differences between thermophiles and mesophiles are in bold.

	NFX	CTX	DTX	TLX	PVX	thermo	TRX II	BCX	THX	TRX I	AKX	ANX	meso	BAX
No residues	191	191	199	194	194	**194**	190	185	190	178	182	181	**184**	207
Hydrogen bonds	191 (1.00)	179 (0.94) 177 (0.93)	200 (1.01) 196 (0.98)	199 (1.03)	196 (1.01)	**192** (0.99)	183 (0.96) 192 (1.01)	180 (0.97)	188 (0.99)	170 (0.94)	177 (0.97)	170 (0.94) 169 (0.93) 170 (0.94) 169 (0.93)	**179** (0.97)	204 (0.99) 203 (0.98)
main-main	116 (0.59)	118 (0.62) 118 (0.62)	121 (0.61) 119 (0.60)	117 (0.60)	115 (0.59)	**117** (0.60)	117 (0.62) 120 (0.63)	109 (0.59)	116 (0.61)	108 (0.61)	110 (0.60)	106 (0.59) 106 (0.59) 106 (0.59) 106 (0.59)	**111** (0.60)	126 (0.61) 127 (0.61)
main-side	28 (0.14)	26 (0.14) 24 (0.13)	39 (0.20) 39 (0.20)	43 (0.22)	43 (0.22)	**36** (0.19)	35 (0.18) 39 (0.21)	35 (0.19)	40 (0.21)	27 (0.15)	29 (0.16)	26 (0.14) 26 (0.14) 26 (0.14) 26 (0.14)	**32** (0.17)	42 (0.20) 42 (0.20)
side-side	**47** (0.25)	35 (0.18) 35 (0.18)	40 (0.20) 38 (0.19)	39 (0.20)	38 (0.20)	**40** (0.21)	31 (0.16) 33 (0.17)	36 (0.19)	37 (0.20)	22 (0.12)	38 (0.21)	38 (0.21) 37 (0.20) 38 (0.21) 37 (0.20)	**34** (0.19)	36 (0.17) 34 (0.16)
Salt bridges	**9** (0.046)	8 (0.042) 7 (0.037)	5 (0.025) 4 (0.020)	7 (0.036)	6 (0.031)	**7** (0.036)	8 (0.042) 8 (0.042)	4 (0.022)	5 (0.026)	2 (0.011)	3 (0.016)	3 (0.017) 3 (0.017) 3 (0.017) 3 (0.017)	**4** (0.022)	**12** (0.058) **12** (0.058)

Table 7. Structure statistics. Protein density = VdW(0 Å)/VdW (1.4 Å); void volume = VdW (1.4 Å) – VdW (0 Å). The largest differences between thermophiles and mesophiles are in bold.

	NFX	CTX	DTX	TLX	PVX	thermo	TRX II	BCX	THX	TRX I	AKX	ANX	meso	BAX
Total number of atoms	1506	1495	1586	1512	1493	1518	1485	1448	1471	1348	1394	1388	1422	1664
Atoms/residue	7.88	7.83	7.97	7.79	7.70	7.83	7.82	7.83	7.74	7.57	7.66	7.67	7.73	**8.04**
ASA (Å^2)	8170	8250	8260	8150	8140	8190	7960	7810	7800	7490	7580	7600	7700	8560
		8140	8340				7900					7600		8490
												7600		
												7600		
ASA/atoms	5.43	5.52	**5.21**	5.39	5.45	5.40	5.36	5.39	5.30	5.56	5.44	5.48	5.42	**5.14**
		5.44	**5.26**				5.32					5.48		**5.10**
												5.48		
												5.48		
Non-polar (%)	50.4	**48.0**	**47.2**	52.0	50.9	49.7	51.3	51.8	53.3	50.5	50.4	50.4	51.3	**48.8**
		48.0	**47.6**				51.6					50.4		49.8
												50.4		
												50.4		
Polar (%)	49.6	**52.0**	**52.8**	48.0	49.1	50.3	48.7	48.2	46.7	49.5	49.6	49.6	48.7	**51.2**
		52.0	**53.4**				48.4					49.6		50.2
												49.6		
												49.6		
Protein density	0.550	0.545	**0.556**	0.544	0.541	0.547	0.549	0.547	0.549	0.543	0.548	0.544	0.547	**0.556**
		0.548	**0.555**				0.550					0.544		**0.559**
												0.545		
												0.544		
Void volume/atoms	10.48	10.69	**10.30**	10.69	10.85	10.59	10.51	10.61	10.50	10.81	10.54	10.71	10.62	**10.26**
		10.57	**10.30**				10.45					10.70		**10.16**
												10.68		
												10.70		

are several minor modifications that correlate with the increased thermostability. Increase in the frequency of Arg is a known determinant in the thermostability, in the same way as the increase in the Thr : Ser ratio. The bulky side chain of Arg offers a possibility for several hydrogen bonds and involvement in salt bridges. A general stabilizing principle appears to be an improved network of interactions, which is reflected in the increased frequency of total number of atoms, charged amino acids and hydrogen bonds. Another feature appears to be the increase in the number of residues in the β-strands. This indicates that the two-layered β-sheet has an important role in the stability of family 11 xylanases. However, the frequencies of amino acids with high β-forming propensity may either decrease (less Val) or increase (higher Thr : Ser ratio). The α-helix region of thermophilic xylanases is stabilized by various strategies including additional hydrogen bonding, salt bridges or disulfide bridges. It is evident that all these changes increase the protein rigidity, which is a property usually associated with enhanced thermostability.

We found also some features that can explain the increased thermostability in specific cases. The highly thermophilic NFX had a great number of side chain–side chain polar interactions and several salt bridges. There could be a trend in xylanase structures that acidophilic xylanases have only few and alkalophilic BAX several salt bridges. It is possible that alkaline and thermal adaptation use partly the same mechanisms for improving the stability. Better packing of the thermostable DTX and the alkalophilic BAX is likely to improve thermostability. Thermophilic CTX and DTX have an increased fractional polar surface, which creates more hydrogen bonds with water.

Several experimental studies have been published on the stability of family 11 xylanases. Two major regions affecting the thermostability are the protein N-terminus and the α-helix region [19,20,47,48,56]. Thermophilic xylanases appear to have a more stable α-helix than their mesophilic counterparts. The mechanism of stabilization at the N-terminus is not so obvious. All thermophilic xylanases have an additional β-strand B1 at the N-terminus, but also mesophilic TRX II and THX contain this β-strand. The N-terminus of NFX is a couple of amino acids longer than that of TRX II, but there is no clear reason why the longer N-terminal might increase the stability. However, the extension of the N-terminus of TRX II has been reported to increase the thermostability [57]. The third region with effect on the thermotolerance seems to be the Ser/Thr surface. The introduction of five arginines into this surface increased the apparent temperature optimum by ≈ 5 °C [39]. Thermophilic CTX has two additional arginines, Arg27 and Arg68, on that surface. Although there are more arginines in thermophilic xylanases, the presence of arginines on Ser/Thr surface does not seem to be a widely used strategy among family 11 xylanases. In conclusion, a number of minor modifications appear to explain the higher stability of thermophilic family 11 xylanases and many thermophilic xylanases have unique features that may increase their stability.

Acknowledgements

We gratefully acknowledge Arja Mäntylä (ROAL, Finland) for providing the enzyme materials. We thank Reetta Kallio-Ratilainen and Johanna Aura for their technical assistance and Dr Xiaoyan Wu for assistance in protein purification. The Academy of Finland and the National Technology Agency of Finland, TEKES, supported this study.

References

1. Viikari, L., Kantelinen, A., Sundquist, J. & Linko, M. (1994) Xylanases in bleaching: from an idea to the industry. *FEMS Microbiol. Rev.* **13**, 335–350.
2. Prade, R.A. (1996) Xylanases: from biology to biotechnology. *Biotechnol. Genet. Eng. Rev.* **13**, 101–131.
3. Bicly, P. (1985) Microbial xylanolytic systems. *Trends Biotechnol.* **3**, 286–290.
4. Henrissat, B. & Davies, G. (1997) Structural and sequence based classification of glycosyl hydrolases. *Curr. Opin. Struct. Biol.* **7**, 637–644.
5. Campbell, R.L., Rose, D.R., Wakarchuk, W.W., To, R., Sung, W. & Yaguchi, M. (1993) *Trichoderma Reesei Cellulases and Other Hydrolases*. (Suominen, P. & Reinikainen, T., eds) pp. 63–72, Foundation for Biochemical and Industrial Fermentation Research, Espoo, Finland.
6. Wakarchuk, W.W., Campbell, R.L., Sung, W.L., Davoodi, J. & Yaguchi, M. (1994) Mutational and crystallographic analyses of the active-site residues of the *Bacillus circulans* xylanase. *Protein Sci.* **3**, 467–475.
7. Sidhu, G., Withers, S.G., Nguyen, N.T., McIntosh, L.P., Ziser, L. & Brayeer, G.D. (1999) *Biochemistry* **38**, 5346–5354.
8. Törrönen, A., Harkki, A. & Rouvinen, J. (1994) Three-dimensional structure of endo-1,4-β-xylanase II from *Trichoderma reesei*: two conformation states in the active site. *EMBO J.* **13**, 2493–2501.
9. Törrönen, A. & Rouvinen, J. (1995) Structural comparison of two major endo-1,4-xylanases from *Trichoderma reesei*. *Biochemistry* **34**, 847–856.
10. Krengel, U. & Dijkstra, B.W. (1996) Three-dimensional structure of endo-1,4-β-xylanase I from *Aspergillus niger*: Molecular basis for its low pH optimum. *J. Mol. Biol.* **263**, 70–78.
11. Gruber, K., Klintschar, G., Hayn, M., Schlacher, A., Steiner, W. & Kratky, C. (1998) Thermophilic xylanase from *Thermomyces lanuginosus*: High-resolution X-ray structure and modeling studies. *Biochemistry* **37**, 13475–13485.
12. Fushinobu, S., Ito, K., Konno, M., Wakagi, T. & Matsuzawa, H. (1998) Crystallographic and mutational analyses of an extremely acidophilic and acid-stable xylanase: Biased distribution of acidic residues and importance of Asp37 for catalysis at low pH. *Protein Eng.* **11**, 1121–1128.
13. Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jorgensen, P.L., Schulein, M., Dupont, C., Davies, G.J. & Wilson, K.S. (1999) Catalysis and specificity in enzymatic glycoside hydrolysis: a 2,5B conformation for the glycosyl-enzyme intermediate revealed by the structure of the *Bacillus agaradhaerens* family 11 xylanase. *Chem. Biol.* **6**, 448–455.
14. Kumar, P.R., Eswaramoorthy, S., Vithayathil, P.J. & Viswamitra, M.A. (2000) The tertiary structure at 1.59 Å resolution and the proposed amino acid sequence of a family-11 xylanase from the thermophilic fungus *Paecilomyces varioti* Bainier. *J. Mol. Biol.* **295**, 581–593.
15. McCarthy, A.A., Morris, D.D., Bergquist, P.L. & Baker, E.N. (2000) Structure of TRX IB, a highly thermostable β-1,4-xylanase from *Dictyoglomus thermophilum* Rt46B.1, at 1.8 Å resolution. *Acta Cryst.* **D56**, 1367–1375.
16. Harris, G.W., Pickersgill, R.W., Connerton, I., Debeire, P., Touzel, J.P., Breton, C. & Perez, S. (1997) Structural basis of the properties of an industrially relevant thermophilic xylanase. *Proteins* **29**, 77–86.
17. Wouters, J., Georis, J., Engher, D., Vandenhaute, J., Dusart, J., Frere, J.M. & Charlier, P. (2001) Crystallographic analysis of family 11 endo-β-1,4-xylanases Xyl1 from *Streptomyces* sp. S38. *Acta Cryst.* **D57**, 1813–1819.
18. Dunlop, R.W., Wang, B., Ball, D., Ruollo, A.B. & Falk, C.J. (1996) Bacterial protein with xylanase activity. Patent US-6200797, Biotech International Limited, Australia.
19. Morris, D.D., Gibbs, M.D., Chin, C.W., Koh, M.H., Wong, K.K.Y., Allison, R.W., Nelson, P.J. & Bergquist, P.L. (1998) Cloning of the *TRX 1B* gene from *Dictyoglomus thermophilum* Rt46B.1 and action of the gene product on Kraft pulp. *Appl. Environ. Microb.* **64**, 1759–1765.
20. Turunen, O., Etuaho, K., Fenel, F., Vehmaanperä, J., Wu, X., Rouvinen, J. & Leisola, M. (2001) Combination of weakly stabilizing mutations with a disulfide-bridge in the α-helix region of *Trichoderma reesei* endo-1,4-β-xylanase II increases the thermal stability through synergism. *J. Biotechnol.* **88**, 37–46.
21. Bailey, M.J., Bicly, P. & Poutanen, K. (1992) Interlaboratory testing of methods for assay of xylanase activity. *J. Biotechnol.* **23**, 257–270.
22. Otwinowski, Z. & Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
23. Navaza, J. (1994) AmoRe: An automated package for molecular replacement. *Acta Cryst.* **A50**, 157–163.
24. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. & Warren, G.L. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Cryst.* **D54**, 905–921.
25. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjelgaard, M. (1991) Improved methods for binding protein models on electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110–119.
26. Perrakis, A., Sixma, T.K., Wilson, K.S. & Lamzin, V.S. (1997) wARP: Improvement end extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models. *Acta Cryst.* **D53**, 448–455.
27. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993) PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
28. Russel, R.B. & Barton, G.J. (1992) Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins: Struct. Funct. Genet.* **14**, 309–323.
29. Kabsch, W. & Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
30. Hubbard, S.J. & Thonton, J.M. (1993) *NACCESS, Computer Program*, available from http://wolf.bms.umist.ac.uk/naccess/
31. Kleywegt, G.J. & Jones, T.A. (1994) Detection, delineation, measurement and display of cavities. *Acta Cryst.* **D50**, 178–185.
32. McDonald, I. & Thornton, J. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.
33. Barlow, D.J. & Thornton, J.M. (1983) Ion-pairs in proteins. *J. Mol. Biol.* **168**, 867–885.
34. Wang, C., Eufemi, M., Turano, C. & Giartosio, A. (1996) Influence of the carbohydrate moiety on the stability of glycoproteins. *Biochemistry* **35**, 7299–7307.

35. Havukainen, R., Törrönen, A., Laitinen, T. & Rouvinen, J. (1996) Covalent binding of three epoxyalkyl xylosides to the active site of endo-1,4-xylanase II from *Trichoderma reesei*. *Biochemistry* **35**, 9617–9624.
36. Menendez-Arias, L. & Argos, P. (1989) Engineering protein thermal stability. Sequence statitistics point to residue substitutions in alpha-helices. *J. Mol. Biol.* **206**, 397–406.
37. Vogt, G., Woell, S. & Argos, P. (1997) Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* **269**, 631–643.
38. Kumar, S., Tsai, C.J. & Nussinov, R. (2000) Factors enhancing protein thermostability. *Protein Eng.* **13**, 179–191.
39. Turunen, O., Vuorio, M., Fenel, F. & Leisola, M. (2002) Engineering of multiple arginines into the Ser/Thr surface of *Trichoderma reesei* endo-1,4-beta-xylanase II increases the thermotolerance and shifts the pH optimum towards alkaline pH. *Protein Eng.* **15**, 141–145.
40. Argos, P., Rossman, M.G., Grau, U.M., Zuber, H., Frank, G. & Tratschin, J.D. (1979) Thermal stability and protein structure. *Biochemistry* **18**, 5698–5703.
41. Facchiano, A.M., Colonna, G. & Ragone, R. (1998) Helix stabilizing factors and stabilization of thermophilic proteins: an X-ray based study. *Protein Eng.* **11**, 753–760.
42. Scholtz, J.M., Qian, H., Robbins, V.H. & Baldwin, R.L. (1993) The energetic of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry* **32**, 9668–9676.
43. Viguera, A.R. & Serrano, L. (1995) Side-chain interactions between sulfur-containing amino acids and phenylalanine in α-helices. *Biochemistry* **34**, 8771–8779.
44. Richardson, J.S. & Richardson, D.C. (1988) *Science* **240**, 1648–1652.
45. Muilu, J., Törrönen, A., Peräkylä, M. & Rouvinen, J. (1998) Functional conformational changes of endo-154-xylanase II from *Trichoderma reesei*: a molecular dynamics study. *Proteins* **31**, 434–444.
46. Daggett, V. & Levitt, M. (1993) Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.* **232**, 600–619.
47. Wakarchuk, W.W., Sung, W.L., Campbell, R.L., Cunningham, A., Watson, D.C. & Yaguchi, M. (1994) Thermostabilization of the *Bacillus circulans* xylanase by the introduction of disulfide bonds. *Protein Eng.* **7**, 1379–1386.
48. Sung, W.L. & Tolan, J.S. (2000) Thermostable xylanases. WO Patent 00/29587.
49. Russell, R.J.M., Ferguson, J.M.C., Hough, D.W., Danson, M.J. & Taylor, G.L. (1997) The crystal structure of citrate synthase from hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution. *Biochemistry* **36**, 9983–9994.
50. DeDecker, B.S., O'Brien, R., Fleming, P.J., Geiger, J.H., Jackson, S.P. & Sigler, P.B. (1996) The crystal stucture of a hyperthermophilic archael TATA-box binding protein. *J. Mol. Biol.* **264**, 1072–1084.
51. Karshikoff, A. & Ladenstein, R. (1998) Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein Eng.* **11**, 867–872.
52. Chen, J., Lu, Z., Sakon, J. & Stites, W.E. (2000) Increasing thermostability of staphylococcal nuclease: implications for the origin of protein thermostability. *J. Mol. Biol.* **303**, 125–130.
53. Chan, M.K., Mukund, S., Kletzin, A., Adams, M.W.W. & Rees, D.C. (1995) Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **267**, 1463–1469.
54. Burley, S.K. & Petsko, G.A. (1985) Aromatic–aromatic interaction: a mechanism of protein structure stabilization. *Science* **229**, 23–28.
55. Kannan, N. & Vishveshwara, S. (2000) Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng.* **13**, 753–761.
56. Georis, J., de Lemos Esteves, F., Lamotte-Brasseur, J., Bougnet, V., Devreese, B., Giannotta, F., Granier, B. & Frère, J.-M. (2000) An additional aromatic interaction improves the thermostability and thermophilicity of a mesophilic family 11 xylanase: Structural basis and molecular study. *Protein Sci.* **9**, 466–475.
57. Sung, W.L., Yaguchi, M. & Ishikawa, K. (1998) Modification of xylanase to improve thermophilicity, alkophilicity and thermostability for pulp bleaching. Patent US-5759840, National Research Council of Canada, Canada.

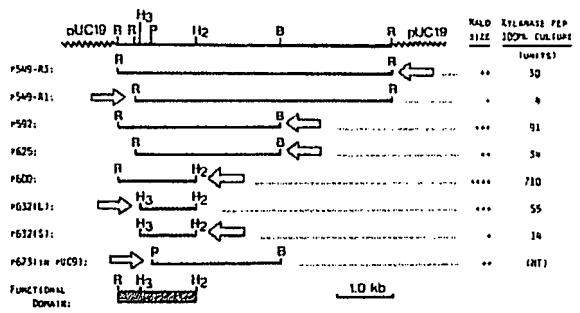


FIG. 1. Mapping of the functional domain of *B. circulans* xylanase gene cloned in pBCX549-R3. The 4.3-kb *B. circulans* DNA insert was subjected to subcloning at various restriction sites (B, *Bam*HI; H_2, *Hin*cII; H_3, *Hin*dIII; P, *Pst*I; R, *Eco*RI). Deleted areas are indicated with dotted lines. Xylanase expression levels obtained with the various subclones are given on the right. Clearing zones extending 1, 2, 3, and 4 mm from colonies are designated +, ++, +++, and ++++, respectively. The location and extent of the xylanase gene functional domain are indicated by the shaded box. The solid box indicates the location of enhancer sequence, and the arrow indicates the direction and location of the LacZ' gene promoter. All subclones carried pUC19 as the vector with the exceptions of pBCX632(L) and pBCX632(S), which carried pTZ18 and pBCX673 which carried pUC9. NT, Not tested.

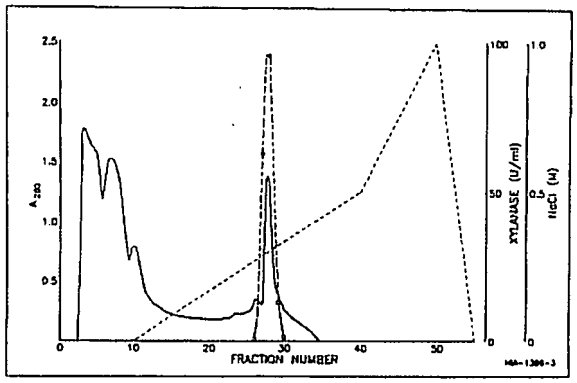


FIG. 2. TSK CM chromatography of intracellular xylanase produced by *E. coli* harboring pBCX549-R3. Symbols: ——, A_{280}; - - -, NaCl; O, xylanase.

macia). Protein bands were visualized by Coomassie brilliant blue staining.

RESULTS

Mapping of the xylanase gene functional domain in pBCX549-R3. The functional domain of the xylanase gene was located by subcloning of the 4.3-kilobase (kb) *B. circulans* DNA insert. The subcloning strategy and the results of the xylanase assays are summarized in Fig. 1. A 1.0-kb *Hin*dIII-*Hin*cII subfragment was found to code for xylanase activity. After 24 h of growth, virtually all of the activity produced by the various subclones was located in the cytoplasmic fraction. However, the amount of activity produced by the different clones varied greatly. Maximum levels were observed with pBCX600, a subclone in which the 0.3-kb upstream sequence was retained and the noncoding downstream sequence was deleted.

Expression levels. The influence of the upstream sequence on expression level is demonstrated by the differences in xylanase production by the two primary clones (pBCX549-R1 and pBCX549-R3) and by the effect of elimination of the sequence in some of the subclones. The difference in expression level between pBCX600 and pBCX632(S) is particularly striking. It was observed that xylanase overproduction by the pBCX600 clone was accompanied by a change in cellular morphology. Many of the cells were up to 10 times longer than normal *E. coli* cells. Deletion of the sequence downstream from the coding region resulted in a significant increase in expression level. The different xylanase levels produced by pBCX632(L) and pBCX632(S) suggest that the orientation of the gene (31) with respect to the β-galactosidase promoter is of some importance.

The effects of the upstream sequence and the β-galactosidase promoter orientation were further investigated by construction of a subclone (pBCX839-3) from pBCX600 in which the orientation of the *Hin*cII-*Hin*dIII fragment was reversed (data not shown). This was accompanied by a 50% reduction in expression level. Clones pBCX839-3 and pBCX632(L) are identical with respect to β-galactosidase promoter and xylanase gene orientation, yet the expression level for pBCX839-3 was seven times that of pBCX632(L). This indicates that the *Eco*RI-*Hin*dIII fragment also enhances expression level when positioned downstream of the xylanase gene.

Subclone pBCX673, which was constructed by insertion of the 2.2-kb *Pst*I-*Bam*HI fragment of pBCX549-R3 into the corresponding sites of pUC9, was halo positive (Fig. 1), while a recombinant constructed in the same manner with pUC19 as the vector was halo negative. The nucleotide sequence near the *Pst*I site of pBCX673 indicated that the first 10 codons of the β-galactosidase gene contained in pUC9 had been inserted immediately upstream of the mature xylanase gene and in the same reading frame: 5'. . . . ATG ACC ATG ATT ACG CCA AGC TTG GCT GCA GCT AGC. . . 3' (*Pst*I site underlined). In effect, the 10-residue N-terminal peptide encoded by the β-galactosidase gene replaced the 28-residue prepeptide of the wild-type xylanase.

Levels of xylanase production by *E. coli* harboring pBCX600 in 2YT medium and 2YT supplemented with larchwood or oat-spelt xylan were similar. Addition of isopropyl-β-D-thiogalactopyranoside was also without effect. Addition of glucose or xylose to the medium resulted in much-reduced enzyme levels. For this reason, the organism was grown in unsupplemented 2YT for xylanase production.

Southern hybridization analyses. Brief autoradiography showed that *Eco*RI digests of *B. circulans* and *B. subtilis* genomic DNA contained a common 4.0-kb band that hybridized with pBCX549-R3 (data not shown). The existence of a common 22-kDa xylanase gene in these organisms was further confirmed by hybridization between pBCX549-R3 and *Bam*HI digests of genomic DNA. Digests of *B. circulans* and *B. subtilis* DNA contained 9- and 15-kb bands, respectively, that hybridized strongly with pBCX549-R3.

Xylanase purification. After 24 h of growth, more than 95% of the activity was located in the intracellular fraction. Fractionation of the intracellular material on a TSK CM column separated the xylanase activity, which adsorbed to the column, from most of the other UV-absorbing activity (Fig. 2). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of the xylanase fraction revealed the presence of small amounts of impurities. The xylanase fraction obtained by ion exchange chromatography was further purified by permeation chromatography on a TSK G2000 column. The activity was associated with the larger of two protein peaks

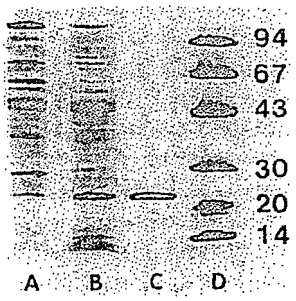


FIG. 3. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of intracellular fraction of *E. coli* harboring pUC19 (A), intracellular fraction of *E. coli* harboring pBCX549-R3 (B), purified xylanase (C), and molecular weight markers (D). Numbers on the right are molecular weights (10^3) of the markers. Approximately 20 μg of intracellular protein and 2 μg of purified xylanase were applied to the gel.

and eluted late in the run (data not shown). The enhanced expression of the enzyme in *E. coli* harboring pBCX600 is evident from the size of the xylanase peak in the TSK CM column chromatogram (Fig. 2) and the intensity of the xylanase band in the intracellular fraction of cells harboring this plasmid (Fig. 3).

Xylanase characteristics. The active fraction obtained by the purification procedure was shown to contain a single band with a molecular weight of approximately 22,000 (Fig. 3). On the basis of sequence data (31), the actual molecular weight of the protein is known to be 20,500. The protein eluted from the gel permeation column at a position much later than would be expected for a molecule of this size. The enzyme had an adsorption maximum at 280 nm and a molar extinction coefficient of 53,000. It apparently bound Coomassie blue poorly, since the Bradford protein assay gave a value that was 1/10 the actual amount, as determined from the weight of lyophilized pure enzyme.

DISCUSSION

It was previously reported (31) that the coding sequence of the pBCX600 insert was virtually identical to that described earlier (18) for the *B. subtilis* gene which codes for 22-kDa xylanase. The expression level data reported in the present article point to the significance of the 0.3-kb upstream fragment, since its deletion resulted in a drastic drop in intracellular xylanase levels. The nucleotide sequence of this fragment has been determined and has been shown to contain a modified promoter sequence and two pairs of tandem repeats resembling some animal virus enhancer elements (1, 31, 32). The *Eco*RI-*Hin*dIII fragment described in this report behaves like eucaryotic transcriptional enhancer elements in that expression levels are influenced by the location of the fragment in positions both upstream and downstream of the xylanase gene. The reason for the increased expression that was observed upon deletion of the sequence downstream from the coding region is unknown. It could be the result of increased copy number of the smaller plasmid and enhanced translation of the xylanase gene in the smaller plasmid. There is a (N) . . . -Ala-Ala-Ser- . . . (C) juncture sequence between the prepeptide and mature peptide (31). This sequence is also found in the pBCX673 subclone, in the genes encoding several extracellular *B. subtilis* proteins (14), and in other xylanases (9, 18). The fragment also contains a Shine-Dalgarno sequence (24) or putative ribosome-binding site. The 3'-flanking sequence of the insert contains a pair of inverted repeats which have been proposed as transcription termination signals (21, 31).

The upstream sequence contained in the clone described here is considerably longer than that reported for other clones of homologous genes from other bacilli. A ribosome-binding site and possible promoter sequences were identified in a 60-base-pair (bp) sequence upstream of the initiation codon in a clone of the *B. pumilus* gene (9). A clone of the *B. subtilis* gene contained a 115-bp sequence upstream from the initiation codon (18). Possible promoter sequences were also identified in this sequence. The upstream sequence in the *B. circulans* clone described here contains 391 bp. A putative promoter sequence is located at 50 to 57 bp upstream of the initiation codon. The two pairs of tandem repeats that are thought to act as enhancer elements (31) are located at 256 to 269 bp and 286 to 302 bp upstream and are not present in the *B. pumilus* or *B. subtilis* clones. Very high expression levels in *E. coli* have been reported for two xylanase genes from two strict anaerobes, *Bacteroides succinogenes* (25) and *Clostridium acetobutylicum* (33). The *C. acetobutylicum* enzyme has a molecular weight of approximately 28,000 (33). The *Bacteroides succinogenes* enzyme has not been characterized. Regulation of gene expression was not studied in either instance.

Commercial application of cloned endoglycanase activities such as use of cellulases and xylanases in the processing of lignocellulosic activities is hampered by low enzyme yields. However, there have now been several reports of high expression levels of these enzymes in *E. coli*. O'Neill et al. (17) obtained overexpression of a *Cellulomonas fimi* cellulase gene in *E. coli* by placing the gene under the control of the leftward promoter of the bacteriophage lambda. Enhanced gene expression was obtained by thermal inactivation of the heat-sensitive lambda *c*I857 repressor. The same strategy has been used to obtain high levels of *Clostridium thermocellum* cellulase expression in *E. coli* (22). *C. thermocellum* cellulase overproduction in *E. coli* has also been reported by Joliff et al. (12). As with the cloned *B. circulans* xylanase gene described here, overexpression was fortuitous. Since the *C. thermocellum* gene was not sequenced, no attempt was made to explain the molecular basis of overproduction.

LITERATURE CITED

1. Banerji, J., S. Rusconi, and W. Schaffner. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell 27:299–308.
2. Biely, P. 1985. Microbial xylanolytic systems. Trends Biotechnol. 3:286–290.
3. Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction method for screening recombinant DNA. Nucleic Acids Res. 7:1513–1522.
4. Bradford, M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein using the principles of protein-dye binding. Anal. Biochem. 72:248–254.
5. Desrochers, M., L. Jurasek, and M. G. Paice. 1981. Production of cellulase, β-glucosidase, and xylanase by *Schizophyllum commune* grown on a cellulose-peptone medium. Dev. Ind. Microbiol. 22:679–684.
6. Fournier, A. R., M. M. Frederick, J. R. Frederick, and P. J. Reilly. 1985. Purification and characterization of endoxylanases from *Aspergillus niger*. An enzyme of pI 3.65. Biotechnol. Bioeng. 27:539–546.
7. Frederick, M. M., J. R. Frederick, A. R Fratzke, and P. J. Reilly. 1981. Purification and characterization of a xylobiose- and xylose-producing endo-xylanase from *Aspergillus niger*.

Carbohydr. Res. 97:87–103.
8. Frederick, M. M., C.-H. Kiang, J. R. Frederick, and P. J. Reilly. 1985. Purification and characterization of endo-xylanases from *Aspergillus niger*. I. Two isoenzymes active on xylan backbones near branch points. Biotechnol. Bioeng. 27:525–532.
9. Fukusaki, E., W. Panbangred, A. Shinmyo, and H. Okada. 1984. The complete nucleotide sequence of the xylanase gene (xynA) of *Bacillus pumilis*. FEBS Lett. 171:197–201.
10. Guo, L.-H., R. C. A. Yang, and R. Wu. 1983. An improved strategy for rapid sequencing of both strands of long DNA molecules cloned in a plasmid. Nucleic Acids Res. 11:5521–5540.
11. Honda, H., T. Kudo, and K. Horikoshi. 1985. Molecular cloning and expression of the xylanase gene of alkalophilic *Bacillus* sp. strain C-125 in *Escherichia coli*. J. Bacteriol. 161:784–785.
12. Joliff, G., P. Beguin, M. Juy, J. Millet, A. Ryter, R. Poljak, and J. P. Aubert. 1986. Isolation, crystallization and properties of a new cellulase of *Clostridium thermocellum* overproduced in *Escherichia coli*. BioTechnology 4:896–900.
13. Jurasek, L., and M. G. Paice. 1988. Xylanase A of *Schizophyllum commune*. Methods Enzymol. 160:659–662.
14. MacKay, R. M., A. Lo, G. Willick, M. Zuker, S. Baird, M. Dove, F. Mornanelli, and V. Seligy. 1986. Structure of a *Bacillus subtilis* endo-β-1,4-glucanase gene. Nucleic Acids Res. 14:9159–9170.
15. MacKenzie, C. R., D. Bilous, and K. G. Johnson. 1984. Purification and characterization of an exoglucanase from *Streptomyces flavogriseus*. Can. J. Microbiol. 30:1171–1178.
16. Meagher, M. M., B. Y. Tao, J. M. Chow, and P. J. Reilly. 1988. Kinetics and subsite mapping of a D-xylobiose- and D-xylose-producing *Aspergillus niger* endo-(1→4)-β-D-xylanase. Carbohydr. Res. 173:273–283.
17. O'Neill, G. P., D. G. Kilburn, R. A. J. Warren, and R. C. Miller, Jr. 1986. Overproduction from a cellulase gene with a high guanosine-plus-cytosine content in *Escherichia coli*. Appl. Environ. Microbiol. 52:737–743.
18. Paice, M., R. Bourbonnais, M. Desrochers, L. Jurasek, and M. Yaguchi. 1986. A xylanase gene from *Bacillus subtilis*: nucleotide sequence and comparison with *B. pumilis* gene. Arch. Microbiol. 144:201–206.
19. Paice, M. G., L. Jurasek, M. R. Carpenter, and L. B. Smillie. 1978. Production, characterization, and partial amino acid sequence of xylanase A from *Schizophyllum commune*. Appl. Environ. Microbiol. 36:802–808.
20. Panbangred, W., E. Fukusaki, E. C. Epifanio, A. Shinmyo, and H. Okada. 1985. Expression of a xylanase gene of *Bacillus pumilis* in *Escherichia coli* and *Bacillus subtilis*. Appl. Microbiol. Biotechnol. 22:259–264.
21. Platt, T. 1981. Termination of transcription and its regulation in the tryptophan operon of *E. coli*. Cell 24:10–23.
22. Schwarz, W. H., S. Schimming, and W. L. Staudenbauer. 1987. High-level expression of *Clostridium thermocellum* cellulase genes in *Escherichia coli*. Appl. Microbiol. Biotechnol. 27:50–56.
23. Shei, J. C., A. R. Fratzke, M. M. Frederick, J. R. Frederick, and P. J. Reilly. 1985. Purification and characterization of endo-xylanases from *Aspergillus niger*. An enzyme of pI 4.5. Biotechnol. Bioeng. 27:533–538.
24. Shine, J., and L. Dalgarno. 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome-binding sites. Proc. Natl. Acad. Sci. USA 71:1342–1346.
25. Sipat, A., K. A. Taylor, R. Y. C. Lo, C. W. Forsberg, and P. J. Krell. 1987. Molecular cloning of a xylanase gene from *Bacteroides succinogenes* and its expression in *Escherichia coli*. Appl. Environ. Microbiol. 53:477–481.
26. Tan, L. U. L., K. K. Y. Wong, and J. N. Saddler. 1985. Functional characteristics of two D-xylanases purified from *Trichoderma harzianum*. Enzyme Microb. Technol. 7:431–436.
27. Tan, L. U. L., K. K. Y. Wong, E. K. C. Yu, and J. N. Saddler. 1985. Purification and characterization of two D-xylanases from *Trichoderma harzianum*. Enzyme Microb. Technol. 7:425–430.
28. Tan, L. U. L., E. K. C. Yu, G. W. Louis-Seize, and J. N. Saddler. 1987. Inexpensive, rapid procedure for bulk purification of cellulase-free β-1,4-D-xylanase of high specific activity. Biotechnol. Bioeng. 30:96–100.
29. Yang, R. C. A., C. R. MacKenzie, D. Bilous, and S. A. Narang. 1989. Identification of two distinct *Bacillus circulans* xylanases by molecular cloning of the genes and expression in *Escherichia coli*. Appl. Environ. Microbiol. 55:568–572.
30. Yang, R. C. A., C. R. MacKenzie, D. Bilous, V. L. Seligy, and S. A. Narang. 1988. Molecular cloning and expression of a xylanase gene from *Bacillus polymyxa* in *Escherichia coli*. Appl. Environ. Microbiol. 54:1023–1029.
31. Yang, R. C. A., C. R. MacKenzie, and S. A. Narang. 1988. Nucleotide sequence of a *Bacillus circulans* xylanase gene. Nucleic Acids Res. 16:7187.
32. Yang, R. C. A., and R. Wu. 1979. Comparative study of papovavirus DNA:BKV(MM), BKV(WT) and SV40. Nucleic Acids Res. 7:651–668.
33. Zappe, H., D. T. Jones, and D. R. Woods. 1987. Cloning and expression of a xylanase gene from *Clostridium acetobutylicum* P262 in *Escherichia coli*. Appl. Microbiol. Biotechnol. 27:57–63.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Mar. 1996, p. 1036–1044
0099-2240/96/$04.00+0

Vol. 62, No. 3

Expression of a *Trichoderma reesei* β-Xylanase Gene (*XYN2*) in *Saccharomyces cerevisiae*

DANIËL C. LA GRANGE, ISAK S. PRETORIUS, AND WILLEM H. VAN ZYL*

Department of Microbiology, University of Stellenbosch, Stellenbosch 7600, South Africa

Received 14 April 1995/Accepted 21 December 1995

The *XYN2* gene encoding the main *Trichoderma reesei* QM 6a endo-β-1,4-xylanase was amplified by PCR from first-strand cDNA synthesized on mRNA isolated from the fungus. The nucleotide sequence of the cDNA fragment was verified to contain a 699-bp open reading frame that encodes a 223-amino-acid propeptide. The *XYN2* gene, located on *URA3*-based multicopy shuttle vectors, was successfully expressed in the yeast *Saccharomyces cerevisiae* under the control of the alcohol dehydrogenase II (*ADH2*) and phosphoglycerate kinase (*PGK1*) gene promoters and terminators, respectively. The 33-amino-acid leader peptide of the Xyn2 β-xylanase was recognized and cleaved at the Kex2-like Lys-Arg residues, enabling the efficient secretion and glycosylation of the heterologous β-xylanase. The molecular mass of the recombinant β-xylanase was estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis to be 27 kDa. The construction of *fur1 ura3 S. cerevisiae* strains allowed for the autoselection of the *URA3*-based *XYN2* shuttle vectors in nonselective complex medium. These autoselective *S. cerevisiae* strains produced 1,200 and 160 nkat of β-xylanase activity per ml under the control of the *ADH2* and *PGK1* promoters in rich medium, respectively. The recombinant enzyme showed highest activity at pH 6 and 60°C and retained more than 90% of its activity after 60 min at 50°C.

Xylan is a major component of the cell walls of monocots and hardwoods, representing up to 35% of the dry weight of these plants (31). This polymer is second only to cellulose in natural abundance and represents a major reserve of fixed carbon in the environment. Unlike cellulose, xylan is a complex polymer consisting of a β-D-1,4-linked xylopyranoside backbone substituted with acetyl, arabinosyl, and glucuronosyl side chains. Hydrolysis of the xylan backbone is catalyzed by endo-β-1,4-xylanases (EC 3.2.1.8) and β-D-xylosidases (EC 3.2.1.37) (5). Endo-β-xylanases act on xylans and xylo-oligosaccharides, producing mainly mixtures of xylo-oligosaccharides. β-D-Xylosidases hydrolyze xylo-oligosaccharides to D-xylose (5). Many bacterial and fungal species are able to utilize xylans as a carbon source. Interest in the enzymology of xylan hydrolysis has recently increased because of the application of β-xylanases in bioblenching (27, 42) and in the food (23) and animal feed (45) industries. Several microbial sources have been investigated for β-xylanase production, and strains of the fungus *Trichoderma* have been shown to secrete large amounts of efficient xylan-degrading enzymes (48).

The yeast *Saccharomyces cerevisiae* can neither utilize nor degrade xylan (18), but it possesses a number of attributes that render it an attractive host for the expression and production of β-xylanases (32). It is a unicellular fungus which provides for posttranslational processing such as endoproteolytic cleavage and glycosylation. *S. cerevisiae* normally secretes few proteins in low abundance, so that a secreted heterologous protein is produced separate from the majority of yeast proteins. Furthermore, it has complete GRAS (generally regarded as safe) status, which allows for its use in the food industry.

Trichoderma reesei is a filamentous mesophilic fungus that is well known for its cellulolytic and xylanolytic enzymatic activities (29). The two major inducible endo-xylanases secreted by this fungus are Xyn1 and Xyn2 (44). They are both relatively small protein molecules with molecular masses of 19 and 21 kDa, respectively. The isoelectric points of Xyn1 and Xyn2 are 5.2 and 9.0, respectively. The pH optima of Xyn1 (pH 3.5 to 4.0) and Xyn2 (pH 4.5 to 5.5) differ, but the proteins produce similar hydrolysis end products (44). Xyn2 represents more than 50% of the total xylanolytic activity of *T. reesei* cultivated on xylan and, together with Xyn1, accounts for more than 90% of the xylan-degrading ability of this fungus (43).

In this paper, we describe the molecular cloning of the *T. reesei XYN2* gene in *S. cerevisiae*. Expression of the *XYN2* gene in *S. cerevisiae* was obtained with the aid of multicopy plasmids, using two different *S. cerevisiae* promoter-terminator expression cassettes derived from the inducible alcohol dehydrogenase II (*ADH2*) gene (35) and the constitutive 3-phosphoglycerate kinase (*PGK1*) gene (15). The enhanced production of recombinant β-xylanase in nonselective complex medium, without the risk of losing the episomal vector, was obtained by disrupting the uracil phosphoribosyltransferase (*FUR1*) gene (19) in the β-xylanase-producing *S. cerevisiae* strains. This step ensured autoselection of the *URA3*-bearing expression plasmids in rich growth medium.

MATERIALS AND METHODS

Strains and media. The genotypes of the microbial strains and plasmids used in the present study are summarized in Table 1. *S. cerevisiae* Y294 was cultivated on either rich YPD medium (1% yeast extract, 2% peptone, 2% glucose) or selective, buffered synthetic complete (SC) medium (2% glucose, 0.67% yeast nitrogen base [Difco] containing amino acid supplements, 20 mM succinate [pH 6.0]), respectively (33). *T. reesei* QM 6a was cultivated in basal medium [0.3% oat spelts xylan (Sigma), 0.4% KH_2PO_4, 1% $(NH_4)_2HPO_4$, 1% peptone, 0.3% yeast extract] (25). Both these organisms were cultured in 1-liter Erlenmeyer flasks containing 100 to 200 ml of medium at 30°C on a rotary shaker at 150 rpm.

Recombinant plasmids were constructed and amplified in *Escherichia coli* JM101 cultivated at 37°C in Luria-Bertani liquid medium or on Luria-Bertani agar (38). Ampicillin for selecting and propagating resistant bacteria was added to a final concentration of 100 μg/ml.

DNA manipulations and plasmid constructions. Standard protocols were followed for DNA manipulations (38). The enzymes for DNA cleavage and ligation and phosphorylated synthetic linkers were purchased from Boehringer-Mann-

* Corresponding author. Mailing address: Department of Microbiology, University of Stellenbosch, Victoria St., Stellenbosch 7600, South Africa. Phone: 27-21-8083160. Fax: 27-21-8083611. Electronic mail address: whvzlp@maties.sun.ac.za.

TABLE 1. Microbial strains and plasmids used in this study

Strain or plasmid	Relevant genotype	Source or reference
Strains		
S. cerevisiae Y294	α *leu2-3,112 ura3-52 his3 trp1-289*	This laboratory
T. reesei QM 6a	Type culture	ATCC 13631
E. coli JM101	r_K^- m_K^+ *supE thiD(lac-proAB)* (F' *traD36 proAB lacI*^{q}Z ΔM15)	38
Plasmids		
YEpC-PADH2a	*bla LEU2 ADH2*$_{PT}$	30
pUC8	*bla tet*	46
pUC19	*bla tet*	49
pBL1	*bla ADH2*$_P$	This laboratory
pBL2	*bla ADH2*$_T$	This laboratory
YEp352	*bla URA3*	14
pDLG1	*bla URA3 ADH2*$_{PT}$	This work
pJC1	*bla URA3 PGK1*$_{PT}$	8
pDLG5	*bla URA3 ADH2*$_P$*-XYN2-ADH2*$_T$	This work
pDLG6	*bla URA3 PGK1*$_P$*-XYN2-PGK1*$_T$	This work
pPE	*bla TRP1 FUR1*	19
pDF1	*bla fur1::LEU2*	This work
YEp13	*bla LEU2*	7

heim and used as recommended by the supplier. Restriction endonuclease-digested DNA was eluted from agarose gels by the method of Benson (4).

The construction of plasmid pDLG1 is summarized in Fig. 1A. A 1.4-kb *Eco*RI-*Sal*I fragment containing the *S. cerevisiae ADH2* promoter region was recovered from plasmid YEpC-PADH2a (30) and ligated to plasmid pUC8, which had been linearized with *Eco*RI and *Sal*I. The unique *Eco*RV site upstream of the *ADH2* promoter region and the *Hin*cII site (also the *Sal*I site) in the pUC8 moiety were subsequently changed to a *Bam*HI and *Bgl*II site, respectively, through the addition of synthetic linkers to create plasmid pBL1. The *ADH2* terminator region was recovered from YEpC-PADH2a as a 0.83-kb *Sma*I-*Bcl*I fragment and cloned into the *Sma*I and *Bam*HI sites of plasmid pUC8. The *Sma*I site was cleaved and changed to a *Bgl*II site through the addition of a synthetic *Bgl*II linker to generate plasmid pBL2. The *ADH2* promoter region was subsequently isolated as a 1.4-kb *Eco*RI-*Bgl*II fragment from plasmid pBL1 and subcloned into the *Eco*RI and *Bgl*II sites of plasmid pBL2. The following synthetic oligodeoxyribonucleotides were annealed to generate a polylinker with the unique restriction sites *Eco*RI, *Bgl*II, and *Xho*I: XBR-1 (5'-CCTCGAGOAG ATCTGGAATTCC-3') and XBR-2 (5'-GGAATTCCAGATCTCCTCGAGG-3'). The *Bgl*II site between the *ADH2* promoter and terminator regions was cleaved, and the protruding ends were filled in with DNA polymerase I (Klenow fragment). The *Eco*RI-*Bgl*II-*Xho*I polylinker was introduced in place of the *Bgl*II site. The 2.1-kb *Bam*HI-*Hin*dIII promoter-terminator fragment was isolated from the plasmid and cloned into the *Bam*HI and *Hin*dIII sites of the yeast-*E. coli* shuttle vector YEp352 (14). The unique *Eco*RI site in YEp352 was cleaved, the protruding ends were filled in with Klenow fragment, and the blunt ends were religated prior to the introduction of the *ADH2* fragment to ensure that the *Eco*RI within the polylinker remains unique. The final *ADH2* yeast expression vector was named pDLG1. Plasmids pDLG1 and pJC1 (8) were used as expression vectors for *XYN2*.

Plasmid pDF1 was constructed as follows. The 2.2-kb *Pst*I-*Eco*RI fragment containing the *S. cerevisiae FUR1* gene was isolated from plasmid pPE (19) and ligated to pUC19, predigested with *Pst*I and *Eco*RI. The *FUR1* gene was cleaved at the unique *Acc*I and *Cla*I sites, the overhanging ends were filled in with DNA polymerase I (Klenow fragment), and a *Sal*I synthetic linker was added to generate a 240-bp *Acc*I-*Cla*I deletion within the *FUR1* open reading frame. The *LEU2* gene was isolated from plasmid YEp13 (7) as a 2.2-kb *Sal*I-*Xho*I fragment and cloned into the *Sal*I site created within the *FUR1* open reading frame to construct a *fur1::LEU2* disruptive allele. The resulting integrating plasmid was designated pDF1.

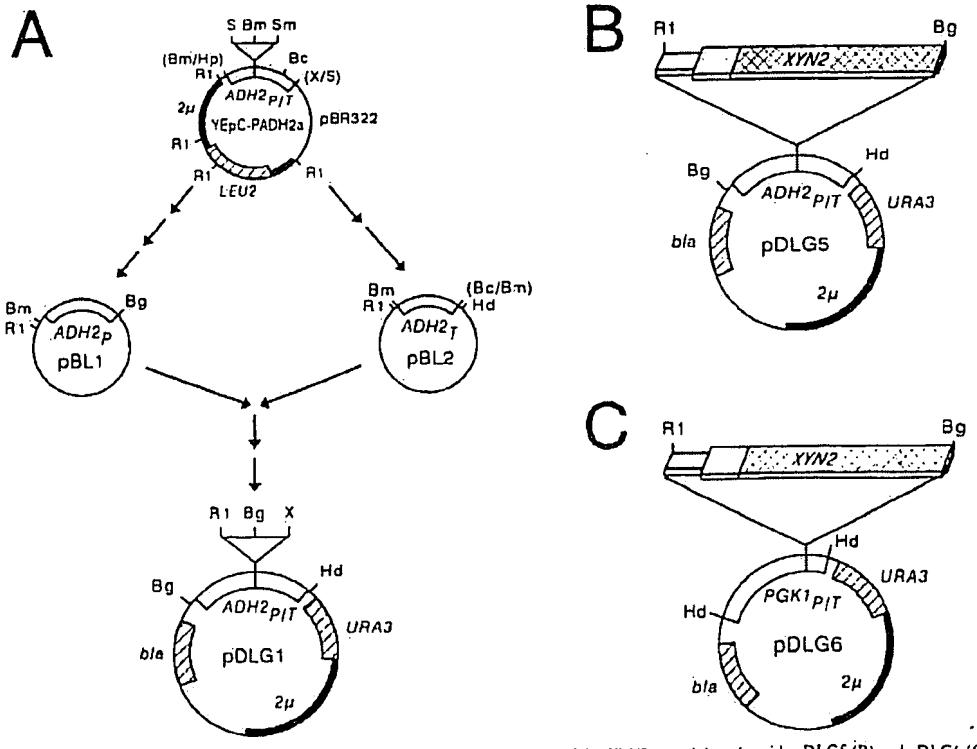


FIG. 1. Schematic summary of the construction of plasmid pDLG1 (A) and plasmid maps of the *XYN2*-containing plasmids pDLG5 (B) and pDLG6 (C). The *XYN2* gene is indicated by cross-hatched boxes, the selectable markers (*LEU2*, *URA3*, and *bla*) are indicated by hatched boxes, the *ADH2* and *PGK1* promoter and terminator sequences are indicated by open boxes, and the 2μm yeast origin of replication and pJR322/pUC8 sequences are indicated by thick and thin lines, respectively. The restriction sites indicated are *Bam*HI (Bm), *Bcl*I (Bc), *Bgl*II (Bg), *Eco*RI (R1), *Hin*dIII (Hd), *Hpa*I (Hp), *Sal*I (S), *Sma*I (Sm), and *Xho*I (X).

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, May 1989, p. 1192–1195
0099-2240/89/051192-04$02.00/0

Vol. 55, No. 5

Hyperexpression of a *Bacillus circulans* Xylanase Gene in *Escherichia coli* and Characterization of the Gene Product†

ROBERT C. A. YANG,* C. ROGER MACKENZIE, DORIS BILOUS, AND SARAN A. NARANG

Division of Biological Sciences, National Research Council of Canada, Ottawa, Ontario, Canada K1A 0R6

Received 19 October 1988/Accepted 6 February 1989

A 4.0-kilobase (kb) fragment of *Bacillus circulans* genomic DNA inserted into pUC19 and encoding endoxylanase activity was subjected to a series of subclonings. A 1.0-kb *Hin*dIII-*Hin*cII subfragment was found to code for xylanase activity. Maximum expression levels were observed with a subclone that contained an additional 0.3-kb sequence upstream from the coding region. Enhancer sequences in the upstream region are thought to be responsible for these high expression levels. Southern hybridization analyses revealed that the cloned gene hybridized with genomic DNA from *Bacillus subtilis* and *Bacillus polymyxa*. Xylanase activity expressed by *Escherichia coli* harboring the cloned gene was located primarily in the intracellular fraction. Levels of up to 7 U/ml or 35 mg/liter were obtained. The protein product was purified by ion exchange and gel permeation chromatography. The xylanase had a molecular weight of 20,500 and an isoelectric point of 9.0.

The structure of xylan, a major component of plant cell walls, is quite complex in that the β-1,4-linked D-xylose backbone is highly substituted with acetyl, arabinosyl, and uronyl groups (2). The enzymatic degradation of xylan to monomeric material requires several distinct activities, but endoxylanase, as the depolymerizing enzyme, is of particular significance. Most biochemical studies have focused on fungal xylanases such as those produced by *Aspergillus niger* (6–8, 16, 23), *Trichoderma harzianum* (26–28), and *Schizophyllum commune* (5, 13, 19). It has been shown that xylanolytic organisms produce multiple forms of xylanase with different physical and kinetic properties (6, 8, 23). It is not known how these enzymes act in concert and to what extent the multiple forms are the result of differential mRNA processing or posttranslational modification.

A 20- to 22-kilodalton (kDa) xylanase has now been identified in xylanase preparations from both fungal and bacterial sources (9, 13, 18, 20, 27, 29, 31). Sequence data for such xylanases from *Bacillus circulans*, *Bacillus subtilis*, *Bacillus pumilus*, *S. commune*, and *T. harzianum* have shown extensive sequence homology (Yaguchi et al., Abstr. Meet. Am. Chem. Soc. 1988, V195, part 1, p. 127). This enzyme has now been identified in several *Bacillus* spp. It is the smallest of at least two distinct xylanases produced by these organisms (11, 29, 30). *B. pumilus* (9, 20), *B. subtilis* (18), and *B. circulans* (29, 31) genes encoding the 20- to 22-kDa enzyme have been cloned and expressed in *Escherichia coli*. On the basis of sequence data and DNA hybridization studies, a high degree of homology has been shown to exist among the 20- to 22-kDa xylanases (9, 18, 31). Expression levels, like those in the *Bacillus* spp. donor strains, have been quite low. This report details the construction of a clone in which the levels of expression of the *B. circulans* xylanase gene are much higher than those previously achieved with cloned xylanase genes or with the *Bacillus* spp. donors.

MATERIALS AND METHODS

Organisms and growth conditions. *E. coli* HB101 harboring recombinant plasmids was grown in 2YT broth as described previously (29). Plasmids pBCX549-R1, pBCX549-R3, and pBCX549-P5 carry *B. circulans* xylanase genes and were constructed as described previously (29). Plasmid pBPX277 contains a *Bacillus polymyxa* xylanase gene (30), and pRH200, kindly provided by M. Paice, harbors a *B. subtilis* xylanase gene (18).

Subcloning of pBCX549-R3. Large preparations of plasmid DNA were isolated by the alkali-sodium dodecyl sulfate method of Birnboim and Doly (3). Small preparations were isolated as described by Guo et al. (10). Plasmid DNA was completely digested with restriction enzymes, and the resulting fragments were separated by electrophoresis in 1.2% low-melting-point agarose gels containing TAE buffer (40 mM Tris hydrochloride [pH 7.8], 20 mM sodium acetate, 2 mM EDTA). Purified subfragments were isolated from the gels (30), ligated with an appropriate vector, and transformed into *E. coli* HB101 (30).

Hybridizations. Southern hybridizations were carried out as described previously (30). *Eco*RI and *Bam*HI digests of genomic DNA from *B. circulans*, *B. polymxya*, and *B. subtilis* were electrophoresed in a 0.8% high-melting-point agarose gel containing TAE buffer and were hybridized with a ^{32}P-labeled pBCX549-R3 probe.

Xylanase assays. Xylanase assays contained 1% oat-spelt xylan and were carried out as described by MacKenzie et al. (15). One unit was defined as the amount of enzyme releasing 1 μmol of reducing sugar per min.

Localization and purification of xylanase. Periplasmic and intracellular fractions were obtained by osmotic shock and sonication procedures as described previously (30). Protein concentrations were estimated by the method described by Bradford (4), with pooled immunoglobulin as the standard. Intracellular xylanase preparations were fractionated by high-pressure liquid chromatography. Samples were applied to a TSK CM-3SW column (7.5 by 150 mm) and eluted at a flow rate of 1 ml/min with a 0 to 1 M NaCl gradient in 10 mM phosphate buffer (pH 6). Fractions containing xylanase were freeze-dried, redissolved, applied to a TSK G2000SW gel permeation column (7.5 by 600 mm), and eluted with 50 mM phosphate buffer (pH 6). Fractions were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis in 10 to 15% polyacrylamide gradient gels in a PhastSystem (Phar-

* Corresponding author.

† National Research Council of Canada publication 30113.

RNA isolation and first-strand cDNA preparation. One liter of *T. reesei* QM 6a culture was prepared in oat spelts basal medium for 48 h at 30°C. The fungal mycelia were harvested through cheesecloth and frozen under liquid nitrogen. The frozen mycelia were ground into a fine powder with a mortar and pestle and suspended in a mixture of 15 ml of phenol and 10 ml of ice-cold STE buffer (100 mM NaCl, 250 mM Tris-HCl [pH 7.2], 10 mM EDTA), and total cellular RNA was isolated as described previously (22). The poly(A)-containing RNA fraction was purified with the aid of an Oligotex-dT mRNA midi kit (Stratagene). First-strand cDNA synthesis was carried out with 100 ng of poly(A) RNA by using a first-strand cDNA synthesis kit (Boehringer Mannheim) as specified by the supplier.

PCR amplification. The *T. reesei XYN2* gene was isolated from a first-strand cDNA mix by PCR with the two oligonucleotides DLG1L (5'-GCATGAATTC GCCAAACCTGAACAACCC-3') and DLG1R (5'-GCATAGATCTCCCTTT AGCTGACGGTGA-3'). These primers were based on the sequence of the *T. reesei XYN2* gene, as published by Törrönen et al. (34). DNA was amplified in 50-μl reaction mixtures (0.25 μM each primer, reaction buffer IV, 1 mM $MgCl_2$, 500 μM each deoxynucleoside triphosphate, 2 μl of template DNA [± 20 ng of first-strand cDNA-mRNA hybrid], 2.5 U of *Taq* polymerase [Advanced Biotechnology]) under mineral oil with a Biometra Trio Thermoblock TB1 (Biometra Biomedizinische Analytik, Gottingen, Germany) (17, 37). Denaturation, annealing, and polymerization were carried out for 1 min at 92°C, 1 min at 55°C, and 1 min at 72°C, respectively, for 29 cycles. The amplified DNA fragment was digested with *Bgl*II and *Eco*RI and ligated to pDLG1 predigested with the same restriction enzymes. The resulting plasmid, pDLG5 (Fig. 1B), was transformed into *S. cerevisiae* Y294 by the dimethyl sulfoxide-lithium acetate method described by Hill et al. (13).

Screening for β-xylanase activity. Transformants were screened for xylan-degrading ability after being plated on SC^{-Ura} medium or SC medium (after disruption of the *FUR1* gene) containing 0.2% of 4-*O*-methyl-D-glucurono-D-xylan–remazol brilliant blue R (RBB)–xylan (Sigma) and 2% galactose as the carbon source (6, 10). β-Xylanase cleaves RBB-xylan into a colorless product.

Subcloning and sequencing of *XYN2*. The β-xylanase gene from pDLG5 was cloned into pUC8, and four deletion subclones were constructed with appropriate restriction enzymes. The nucleotide sequence of the *XYN2* gene was determined by sequencing both strands of the cloned cDNA inserts by the radioactive, dideoxy chain termination method (39) with T7 DNA polymerase (Pharmacia) as specified by the manufacturer. The sequence data obtained were analyzed with the GENEPRO software package (Hoefer Scientific Instruments, San Francisco, Calif.).

Southern and Northern hybridizations. Total DNA was isolated from *S. cerevisiae* strains (16), digested with *Nsi*I and *Nco*I, separated on a 1% agarose gel, and blotted to a Hybond-N membrane (Amersham International). Southern hybridizations were carried out by the method described by Southern (41), modified as described by Sambrook et al. (38). Northern (RNA) hybridizations were carried out as described by Sambrook et al. (38). For both Southern and Northern hybridizations, single-stranded DNA fragments were labelled with a random-primed DNA-labelling kit (Boehringer Mannheim) involving [α-^{32}P] ATP as specified by the manufacturer.

Protein preparation and gel electrophoresis. Protein preparations were performed on ice. The supernatant of a 200-ml culture of the recombinant *S. cerevisiae* strain was separated from the cellular mass by centrifugation for 5 min at 4,000 × *g*. The supernatant was filtered and concentrated in a Diaflo Ultrafilter PM10 concentrator (Amicon Division of W. R. Grace and Co., Danvers, Mass.) and subsequently precipitated with 2 volumes of ice-cold acetone for 1 h. The intracellular protein fraction was extracted after the cells were washed twice in 5 ml of extraction buffer [200 mM Tris (pH 8.0), 400 mM $(NH_4)_2SO_4$, 10 mM $MgCl_2$, 1 mM EDTA, 10% glycerol] and was ultimately suspended in 0.4 ml of the same buffer. Cells were ruptured with ca. 0.2 to 0.3 ml of ice-cold acid-washed glass beads (0.45 μm diameter) in a 2-ml microcentrifuge tube by vortexing six times for 30 s with 60-s intervals on ice. Glass beads and cell debris were removed by centrifugation at 13,500 × *g* for 10 min. The supernatant was subsequently precipitated by the same procedure as with the extracellular fraction. Both protein fractions were resuspended in 0.1 ml of HEPES buffer (20 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 8.0], 5 mM EDTA [pH 8.0], 7 mM β-mercaptoethanol) after centrifugation at 13,500 × *g* for 10 min. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on 15% (wt/vol) polyacrylamide was performed by the method of Laemmli (21). Intracellular and extracellular protein fractions were boiled for 3 min and applied to the gel. Proteins were visualized by Coomassie brilliant blue staining (38).

Protein deglycosylation. The supernatant of a 200-ml culture of the recombinant *S. cerevisiae* was separated from the cellular mass by centrifugation for 5 min at 4,000 × *g* and filtration through a 0.22-μm-pore-size membrane. The recombinant protein of 27 kDa (Xyn2 protein) was separated from the other proteins secreted by *S. cerevisiae* (larger than 33 kDa) by using a Diaflo Ultrafilter PM30 concentrator (Amicon Division of W. R. Grace and Co.) and subsequently precipitated from the filtrate with 2 volumes of acetone for 1 h. The protein fraction was collected by centrifugation at 13,500 × *g* for 90 min at 4°C and resuspended in 0.25 ml of endoglycosidase F incubation buffer (50 mM sodium acetate buffer [pH 5], 20 mM EDTA, 1% 2-mercaptoethanol). The deglycosylation reaction was initiated by adding 0.2 U of endoglycosidase F (Sigma) to 40 μl of denatured protein (boiled for 5 min at 100°C). The reaction mixture was incubated at 37°C for 12 h, after which the sample was denatured again and 0.2 U of fresh endoglycosidase F was added. After a further 16 h of incubation, the reaction mixture was denatured again and SDS-PAGE on 15% (wt/vol) polyacrylamide was performed (21). Untreated Xyn2 protein was treated in the same manner, but no endoglycosidase F was added.

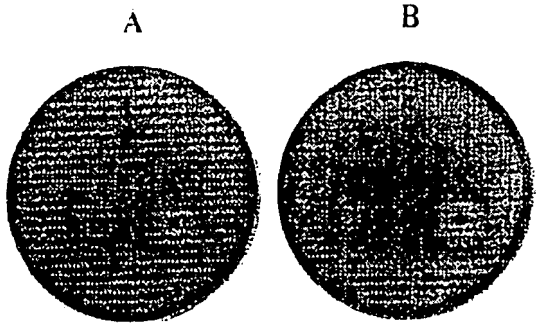


FIG. 2. Recombinant β-xylanase-producing *S. cerevisiae* strains. The colonies at the top of plates A and B are *S. cerevisiae* Y294(*fur1::LEU2* pDLG1) (colonies 1 and 4), those at the bottom left are *S. cerevisiae* Y294(*fur1::LEU2* pDLG5) (colonies 2 and 5), and those at the bottom right are *S. cerevisiae* Y294(*fur1::LEU2* pDLG6) (colonies 3 and 6). Both plates contain SC medium and 0.2% RBB-xylan supplemented with 2% glucose (A) or 2% galactose (B) as the sole carbon source. Colonies degrading RBB-xylan are surrounded by pale clearing zones. The plates were photographed after 30 h of incubation at 30°C.

Enzyme activity assays. Endo-β-1,4-xylanase activity was assayed by the method described by Bailey et al. (1) with 1% birchwood glucuronoxylan (Sigma) as the substrate at 60°C. Appropriate dilutions of the cell-free culture solution in 50 mM sodium citrate buffer (pH 6.0) were used as the enzyme source. The amount of released sugar was determined by the dinitrosalicylic acid method described by Miller et al. (26). Thermostability was tested by heating enzyme samples for different times at various temperatures, and the activity was assayed at 60°C for 5 min as described above. Assays at different pH values were performed as described above, except that the buffers used were 50 mM citrate (pH 3.0), 50 mM citrate phosphate (pH 4.0 to 7.0), and 50 mM phosphate (pH 8.0), respectively (11).

RESULTS

Cloning of the *T. reesei XYN2* gene. The *T. reesei XYN2* gene was isolated from first-strand cDNA prepared from *T. reesei* by using sequence-specific PCR primers. The PCR product was cloned into pDLG1 (creating pDLG5 [Fig. 1B]) under the control of the inducible *ADH2* promoter and transformed into *S. cerevisiae* Y294. β-Xylanase-producing yeast colonies were identified by plating on selective SC medium containing RBB-xylan. The β-xylanase gene was isolated from pDLG5 and cloned into pJC1 (creating pDLG6 [Fig. 1C]) under the control of the constitutive *PGK1* promoter. This plasmid was also introduced into *S. cerevisiae* Y294. The *FUR1* gene of the above-mentioned yeast strains was disrupted to create strains *S. cerevisiae* Y294(*fur1::LEU2* pDLG5) and *S. cerevisiae* Y294 (*fur1::LEU2* pDLG6), respectively. Both these recombinant yeast strains, together with *S. cerevisiae* Y294(*fur1::LEU2* pDLG1) as a control, were plated on SC medium containing 2% glucose or 2% galactose as the carbon source, respectively (Fig. 2A and B). *S. cerevisiae* Y294(*fur1::LEU2* pDLG5) showed smaller clearing zones, because the *ADH2* promoter is strongly repressed by 2% glucose but less so by 2% galactose.

Nucleotide sequence of the *XYN2* gene. The nucleotide and deduced amino acid sequences of *T. reesei XYN2* are presented in Fig. 3. A 780-bp DNA fragment encoding a 223-amino-acid propeptide has been cloned and expressed in *S. cerevisiae*. The 108-bp intron present in the *T. reesei* genomic *XYN2* gene was eliminated through the first-strand cDNA step. The hydropho-

```
                                                                -104  aattcgccaaaacc  -91

-90  tgaacaaccccagcacctgaacagtcatacaaccccctccaagcccaaaagacacaacaactcctactagccgaagcaagaagacatcaac   -1

  1  ATGGTCTCCTTCACCTCCCTCCTCGCCGGCGTCGCCGCCATCTCGGGCGTCTTGGCCGCTCCCGCCGCCGAGGTCGAACCCGTGGCTGTG    90
  1  M  V  S  F  T  S  L  L  A  G  V  A  A  I  S  G  V  L  A  A  P  A  A  E  V  E  P  V  A  V    30

 91  GAGAAGCGCCAGACGATTCAGCCCGGCACGGGCTACAACAACGGCTACTTCCACTCGTACTGGAACGATGGCCACGGCGGCGTGACGTAC   180
 31  E  K  R  Q  T  I  Q  P  G  T  G  Y  N  N  G  Y  F  H  S  Y  W  N  D  G  H  G  G  V  T  Y    60

181  ACCAATGGTCCCGGCGGGCAGTTCTCCGTCAACTGGTCCAACTCGGGCAACTTTGTCGGCGGCAAGGGATGGCAGCCCGGCACCAAGAAC   270
 61  T  N  G  P  G  G  Q  F  S  V  N  W  S  N  S  G  N  F  V  G  G  K  G  W  Q  P  G  T  K  N    90
                                   ↑

271  AAGGTCATCAACTTCTCGGGCAGCTACAACCCCAACGGCAACAGCTACCTCTCCGTGTACGGCTGGTCCCGCAACCCCCTGATCGAGTAC   360
 91  K  V  I  N  F  S  G  S  Y  N  P  N  G  N  S  Y  L  S  V  Y  G  W  S  R  N  P  L  I  E  Y   120
            ↑

361  TACATCGTCGGGAACTTTGGCACCTACAACCCGTCCACGGGCGCCACCAAGCTGGGCGAGGTCACCTCCGACGGCAGCGTCTACGACATT   450
121  Y  I  V  G  N  F  G  T  Y  N  P  S  T  G  A  T  K  L  G  E  V  T  S  D  G  S  V  Y  D  I   150
                        ↑

451  TACCGCACGCAGCGCGTCAACCAGCCGTCCATCATCGGCACCGCCACCTTTTACCAGTACTGGTCCGTCCGCCGCAACCACCGCTCGAGC   540
151  Y  R  T  Q  R  V  N  Q  P  S  I  I  G  T  A  T  F  Y  Q  Y  W  S  V  R  R  N  H  R  S  S   180

541  GGCTCCGTCAACACGGCGAACCACTTCAACGCGTGGGCTCAGCAAGGCCTGACGCTCGGGACGATGGATTACCAGATTGTTGCCGTGGAG   630
181  G  S  V  N  T  A  N  H  F  N  A  W  A  Q  Q  G  L  T  L  G  T  M  D  Y  Q  I  V  A  V  E   210

631  GGTTACTTTAGCTCTGGCTCTGCTTCCATCACCGTCAGCTAAagggagatct  682
211  G  Y  F  S  S  G  S  A  S  I  T  V  S  *              223
```

FIG. 3. Nucleotide and deduced amino acid sequences of the *XYN2* gene from *T. reesei* QM 6a. The hydrophobic leader sequence is underlined, the dibasic residues (Lys-Arg) are double underlined, and the PCR primers sequences used are indicatec in boldface type. Residues marked with arrows are potential targets for N glycosylation. The nucleotide sequence data are available from GenBank nucleotide sequence databases (accession number U24191).

bic leader of the recombinant propeptide produced by *S. cerevisiae* is presumably proteolytically processed by the KEX2 protease, producing a 190-amino-acid mature protein (Fig. 3).

Identification and confirmation of *fur1* disruptions. The *FUR1* gene of *S. cerevisiae* Y294 strains was disrupted with the *fur1*::*LEU2* allele to generate *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG1), *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5), and *S. cerevisiae* Y294 (*fur1*::*LEU2* pDLG6), respectively. The *fur1*::*LEU2* disruptive allele was isolated as a 3.27-kb *Nco*I-*Nsi*I linear DNA fragment for subsequent transformation of *S. cerevisiae* strains to replace the chromosomal *FUR1* gene through gene replacement (34) with the *fur1*::*LEU2* allele (Fig. 4A). Leu$^-$ yeast transformants were selected for on SC^{-Leu} medium. To confirm the *FUR1* gene disruption, *Nsi*I-*Nco*I-digested genomic DNA from *S. cerevisiae* Y294, *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5), and *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6) were subjected to Southern hybridization analysis with a 1.33-kb *Nsi*I-*Nco*I *FUR1* fragment from pPE as an α-^{32}P-labelled probe. In *S. cerevisiae* Y294, a DNA fragment of 1.33 kb was radioactively highlighted, and in *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5) and *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6), DNA fragments of 3.27 kb were highlighted, corresponding to the *FUR1* and *fur1*::*LEU2* alleles, respectively (Fig. 4B). *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG1) without the β-xylanase gene was analyzed in the same way (data not shown).

Northern blot hybridization and SDS-PAGE analysis of cloned proteins. Northern analysis of the mRNA produced by the recombinant *S. cerevisiae* strains showed that high levels of *XYN2* mRNA were produced by *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5) when cultivated for 48 h in YPD with 2% galactose as the carbon source (Fig. 5A). Much lower levels of mRNA were present in *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6) cultivated for 48 h in YPD with 2% glucose as the carbon source. On the SDS-polyacrylamide gel, a new protein band with an estimated molecular mass of 27 kDa was visible for *S. cerevisiae* Y294 (*fur1*::*LEU2* pDLG5) (the more active clone), as well as for *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6). The high level of *XYN2*

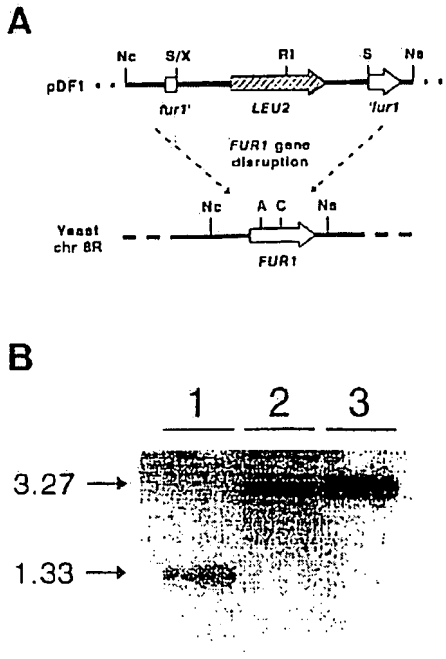


FIG. 4. (A) Schematic representation showing the disruption of the *FUR1* gene on the genome of β-xylanase-producing *S. cerevisiae* strains. The enzymes used were *Acc*I (A), *Cla*I (C), *Eco*RI (RI), *Nco*I (Nc), *Nsi*I (Ns), *Sal*I (S), and *Xho*I (X). (B) Southern blot analysis showing the genomic disruption of *FUR1*. Lanes: 1, *S. cerevisiae* Y294 with the wild-type *FUR1* gene; 2, *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5); 3, *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6). The genomic DNA was hybridized to an α-^{32}P-labelled 1.33-kb *FUR1* DNA fragment.

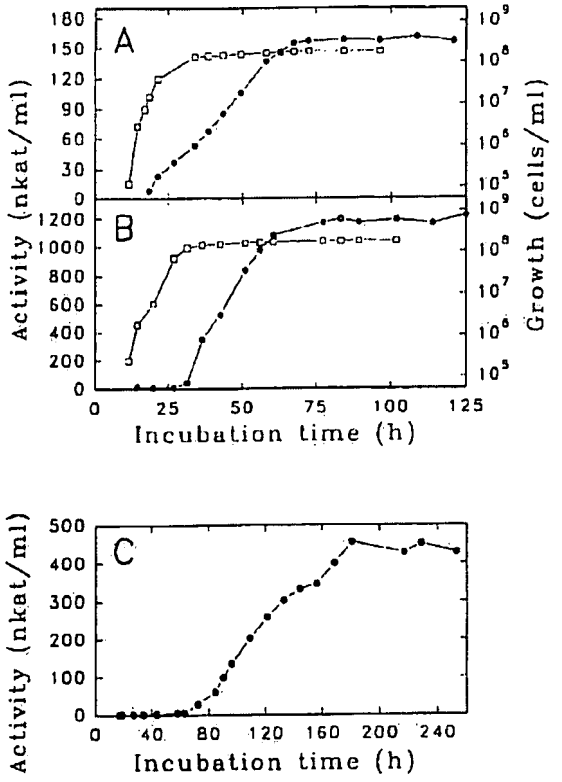


FIG. 6. Time course of extracellular β-xylanase activity produced by *S. cerevisiae* Y294(*furl::LEU2* pDLG6) (A), *S. cerevisiae* Y294(*furl::LEU2* pDLG5) (B), and *T. reesei* (C) in rich medium. The β-xylanase activities were assayed by the method of Bailey et al. (1), as described in the text. Yeast cell counts were determined with the aid of a hemocytometer.

March 1995). The nucleotide sequence of the cDNA fragment containing the *T. reesei* QM 6a *XYN2* gene (Fig. 3) is 99% identical to that reported by Törrönen et al. for the *T. reesei* Rut C-30 *XYN2* DNA sequence (GenBank accession number S51973) (44), differing for only six base pair substitutions and three base pair insertions. The three C-G insertions are at nucleotides 29, 55, and 59 of the *T. reesei* QM 6a *XYN2* DNA sequence, resulting in amino acids 10 to 19 in the proprotein being different from (and more hydrophobic than) those in the proprotein encoded by the *T. reesei* Rut C-30 *XYN2* gene. The cDNA sequence of the *T. reesei* QM 6a *XYN2* gene is identical to the corresponding nucleotide sequence of *T. reesei* VTT-D-79125 (GenBank accession number S67387) reported by Saarelainen et al. (36), except for three base pair substitutions that occurred within the coding region of the *XYN2* gene. These three base pair substitutions also correspond to similar base pair substitutions when compared with the *T. reesei* Rut C-30 *XYN2* nucleotide sequence, suggesting that they are misincorporations by the *Taq* DNA polymerase during the PCR amplification of the *T. reesei* QM 6a *XYN2* cDNA fragment (12). The *XYN2* nucleotide sequences of both *T. reesei* QM 6a and *T. reesei* VTT-D-79125 are probably identical if misincorporation during the PCR amplification is taken into account.

Both *T. reesei* Rut C-30 and *T. reesei* VTT-D-79125 are mutants isolated in mutant selection programs at the U.S. Army Natick Laboratories, Natick, Mass. (24); Rutgers University, Rutgers, N.Y. (40); and VTT-D Technical Research Center, Espoo, Finland (3), that started with *T. reesei* QM 6a as the wild-type strain (Fig. 8). Our sequence data suggest that no mutations were introduced into the *XYN2* gene of *T. reesei* VTT-D-79125 during the three γ-irradiation steps, four *N*-nitro-*N'*-nitro-*N*-nitrosoguanidine steps, and one diethyl sulfate mutagenesis step that were performed prior to its isolation. However, the *XYN2* gene of *T. reesei* Rut-C-30 most probably acquired the three C-G deletion mutations during any of the UV radiation or two *N*-nitro-*N'*-nitro-*N*-nitrosoguanidine mutagenesis steps used in the mutant selection program at Rutgers University, resulting in a less hydrophobic proprotein when compared with *T. reesei* QM 6a and *T. reesei* VTT-D-79125.

The *XYN2* gene was inserted between a yeast promoter and transcription terminator on multicopy episomal plasmids to achieve high levels of gene expression (32). For this purpose, we used the *S. cerevisiae ADH2* and *PGK1* promoter-terminator cassettes. Transcription of *ADH2* is almost undetectable when *S. cerevisiae* is grown on fermentable sugars such as glucose or galactose but is derepressed to a level representing about 1% of total soluble cellular protein when the yeast is grown on nonfermentable carbon sources or fermentable sugars at concentrations lower than 1% (32). The level of *XYN2* expression under the control of the *ADH2* promoter was higher when the yeast transformants were grown on galactose instead of glucose as the carbon source (Fig. 2). The *PGK1* promoter is a strong constitutive promoter which can be induced to a level of expression that constitutes 4 to 10% of the total soluble protein depending on the growth conditions (15). The difference between the two promoters used can clearly be seen in Fig. 2. The *ADH2* promoter is strongly repressed when the cells are plated on glucose, resulting in a large difference in the size of the clearing zones on the plate containing glucose.

Translation in yeasts can be modulated at the level of initiation by four aspects of mRNA structure: (i) the primary sequence or context surrounding the AUG codon; (ii) the position of the AUG codon, i.e., whether it is first; (iii) secondary structure both upstream and downstream of the AUG codon; and (iv) leader length (20). The leader sequence of the *T. reesei XYN2* gene is similar to that of most *S. cerevisiae* genes (20). Therefore, initiation of translation of *XYN2* would be effective in *S. cerevisiae*. This degree of similarity between *S. cerevisiae* and *T. reesei* might be because both these organisms belong to the subdivision *Ascomycotina*.

High-level secretion of heterologous gene products in *S. cerevisiae* is mediated by the hydrophobic N-terminal extension of the polypeptide, the signal/leader sequence. The leader peptide is responsible for translocation of the protein to the endoplasmic reticulum, from where the protein is transported through specialized secretory organelles, modified, processed, and often glycosylated prior to its release by the plasma membrane into the culture medium. The *T. reesei* QM 6a *XYN2* leader peptide has all the characteristics required for processing in *S. cerevisiae*. It is hydrophobic and possesses the dibasic residues Lys-Arg (Fig. 3), which are presumably recognized and cleaved on the carboxyl side by the KEX2 protease of *S. cerevisiae* (32).

However, the Xyn2 β-xylanase secreted by *S. cerevisiae* has a different molecular mass from that of the *T. reesei* enzyme (Table 2). The mature β-xylanase produced by *T. reesei* has a molecular mass of 21 kDa as deduced from the amino acid sequence. The molecular mass determined by SDS-PAGE is almost the same, indicating a virtual absence of carbohydrates (44). The molecular mass of the enzyme secreted by *S. cerevisiae* is 27 kDa as estimated by SDS-PAGE. There is thus a

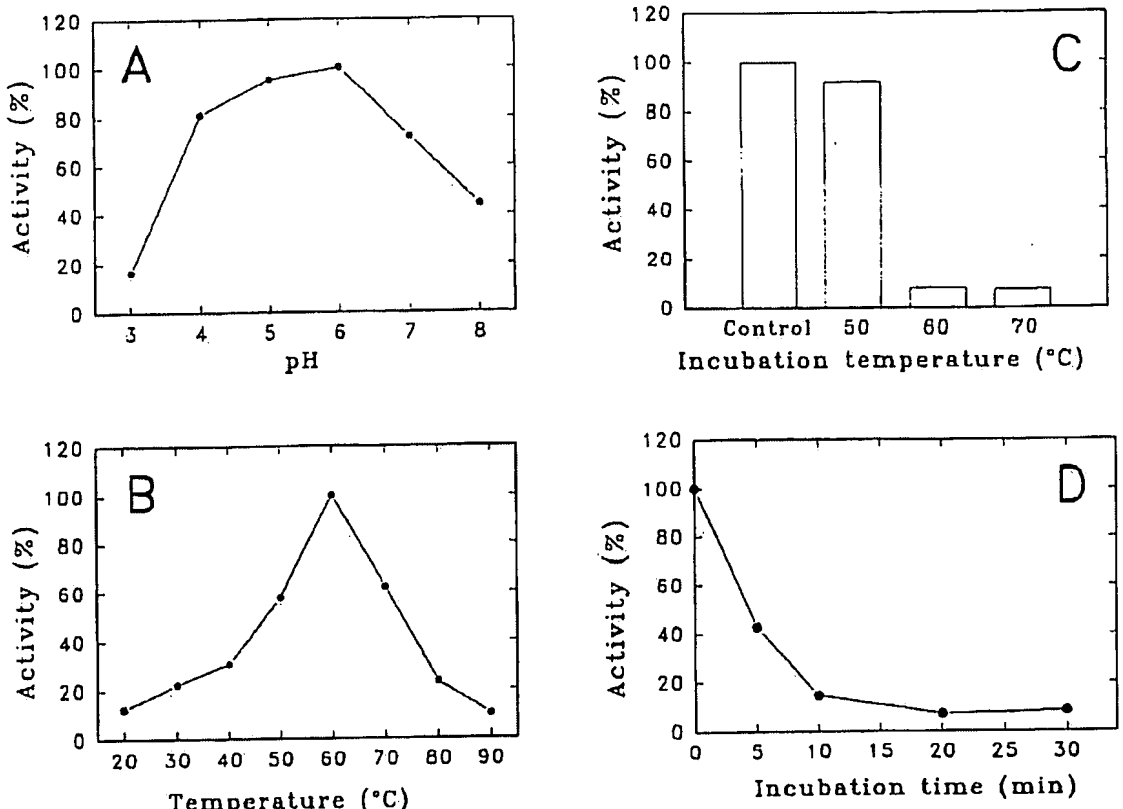


FIG. 7. Effect of pH (A) and temperature (B) on the activity of Xyn2. The highest activity was measured at pH 6 and at 60°C, respectively. The buffers used in the enzyme reactions were 50 mM citrate (pH 3), 50 mM citrate phosphate (pH 4.0 to 7.0), and 50 mM phosphate (pH 8.0). (C) The temperature stability of Xyn2 was determined after preincubating the enzyme in the absence of the substrate for 30 min at 50, 60, and 70°C. (D) The thermostability of Xyn2 at 60°C was determined by preincubating the enzyme at this temperature in the absence of the substrate for 5, 10, 20, and 30 min before determining its activity. The β-xylanase activity prior to the preincubations at different temperatures was taken as 100%.

6-kDa difference in the molecular masses of the β-xylanases secreted by *S. cerevisiae* and *T. reesei*. This is caused by N-glycosylation of the β-xylanase secreted by *S. cerevisiae* (32), because *S. cerevisiae* tends to "hyperglycosylate" heterologous proteins. Treatment of the recombinant Xyn2 protein with endoglycosidase F generated a new protein species with a molecular mass of 21 kDa (Fig. 5C), which corresponds to that of native β-xylanase produced by *T. reesei*. However, this large, glycosylated protein (22% sugar content) was efficiently secreted and passed through the yeast cell wall into the culture medium without any effect on the growth rate. The pH and temperature optima of the recombinant β-xylanase produced by *S. cerevisiae* compare well with those of the native enzyme produced by *T. reesei* (Table 2). Similar hyperglycosylation was observed with the cellobiohydrolases (CBHI and CBHII) of *T. reesei* when expressed in *S. cerevisiae* (28).

The *XYN2* gene is expressed from an episomal plasmid, and the yeast must be kept under selective conditions to ensure vector stability. However, the use of selective synthetic medium is not ideal for the production of high levels of heterologous proteins. We therefore genetically altered the recombinant yeast strains to allow autoselection for the episomal plasmids. The *FUR1* gene of *S. cerevisiae* encodes uracil phosphoribosyltransferase, which catalyzes the conversion of uracil into uridine 5′-phosphate in the pyrimidine salvage pathway (19). If this gene were disrupted, *S. cerevisiae* Y294 would not be able to utilize uracil from the extracellular medium and would therefore not be viable, unless it possessed a complementing functional *URA3* gene to synthesize uridine 5′-phosphate de novo. In this case, the *URA3* gene is the yeast selectable marker on the YEp352-based vectors used for the expression of *XYN2*. After the *FUR1* gene in the β-xylanase-producing *S. cerevisiae* strains was disrupted, these strains could be cultured in YPD medium or any other complete synthetic medium without the risk of losing the episomal plasmid. *S. cerevisiae* could grow to much higher cell densities when cultivated on YPD medium; therefore, β-xylanase activity was considerably higher. In *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5) and *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6), β-xylanase activity increased 24- and 3-fold, respectively, when the cells were cultured on YPD medium instead of selective SC medium (data not shown). The time needed for β-xylanase production to start also decreased substantially.

The highest total β-xylanase activity obtained in shake flasks for the hyperproducing mutant *T. reesei* VTT-D-86271 (Rut C-30) was 5,400 nkat/ml (2). The total activity obtained for the most active recombinant *S. cerevisiae* strain (1,487 nkat/ml) compares well with that of *T. reesei* Rut C-30 if one takes into

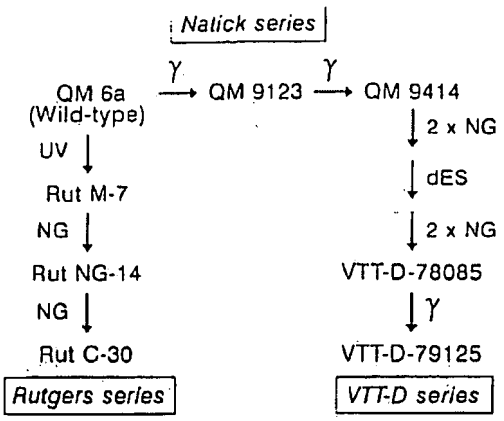


FIG. 8. Lineage of *T. reesei* Rut C-30 and VTT-D-79125 mutants, derived from the wild-type *T. reesei* QM 6a. Mutagenic agents: γ, gamma irradiation; NG, *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine; dES, diethyl sulfate.

account that the *T. reesei* culture supernatant contains the complete battery of enzymes involved in xylan degradation. These include β-xylosidases that degrade the short xylo-oligosaccharides produced by the β-xylanases and debranching enzymes like α-glucuronidases. The cumulative and synergistic action of all these enzymes would therefore lead to larger amounts of reducing sugars being produced compared with a culture with only β-xylanase activity.

Although β-xylanase production by recombinant organisms, such as *S. cerevisiae*, may not be as high as that in organisms such as the above-mentioned *T. reesei* strain, it might still be worthwhile to consider its use for industrial enzyme production. First, *S. cerevisiae* cannot degrade cellulose, xylan, or any of the other polymers in wood; therefore, β-xylanases produced by this organism are pure and completely free of any contaminating cellulases. This is especially important for use in the paper industry. Currently, commercial β-xylanase preparations marketed for pulp treatment include Pulpzyme HA from *T. reesei* (47). This is a relatively crude β-xylanase preparation with residual cellulolytic activity and requires careful control of process parameters to avoid damage to fibers. Second, *S. cerevisiae* can be cultivated on a variety of relatively inexpensive culture media without the need for xylan to induce β-xylanase production. Furthermore, industrial-scale fermentation technology for *S. cerevisiae* is well established.

Future research in our laboratory will be directed toward the development of a more effective expression system for the production of β-xylanases in *S. cerevisiae*. The addition of a β-xylosidase gene and/or gene(s) encoding debranching enzyme(s) will also be considered.

TABLE 2. Molecular characteristics of Xyn2 produced by *T. reesei* and a recombinant Xyn2 produced by *S. cerevisiae*

Property	Value for Xyn2 from:		Reference(s)[a]
	T. reesei	*S. cerevisiae*	
Mol mass (kDa)	20–21	27.5	42, 44
Optimum pH	5	6	42, 44
Optimum temp (°C)	56–60[b]	60	9
Temp stability (60 min) (°C)	NA[c]	50[d]	
No. of N-glycosylation sites	3	3	44
Glycosylation	No	Yes	44

[a] References for *T. reesei* data only.
[b] Temperature optimum of β-xylanases produced by *T. reesei*, not only Xyn2.
[c] NA, not available.
[d] More than 90% of activity remained.

ACKNOWLEDGMENTS

We acknowledge Elizabeth J. Lodolo for the construction of plasmids pBL1 and pBL2.

This work was supported by a grant from National Chemical Products (NCP) (Pty) Ltd.

REFERENCES

1. Bailey, M. J., P. Biely, and K. Poutanen. 1992. Interlaboratory testing of methods for assay of xylanase activity. J. Biotechnol. 23:257–270.
2. Bailey, M. J., J. Buchert, and L. Viikari. 1993. Effect of pH on production of xylanase by *Trichoderma reesei* on xylan- and cellulose-based media. Appl. Microbiol. Biotechnol. 40:224–229.
3. Bailey, M. J., and K. M. H. Nevalainen. 1981. Induction, isolation and testing of stable *Trichoderma reesei* mutants with improved production of solubilizing cellulose. Enzyme Microb. Technol. 3:153–157.
4. Benson, S. A. 1984. A rapid procedure of isolation of DNA fragments from agarose gels. BioTechniques 2:66–67.
5. Biely, P. 1985. Microbial xylanolytic systems: Trends Biotechnol. 3:286–290.
6. Biely, P., D. Mislovicova, and R. Toman. 1988. Remazol brilliant blue-xylan: a soluble chromogenic substrate for xylanases. Methods Enzymol. 160:536–541.
7. Broach, J. R., J. N. Strathern, and J. B. Hicks. 1979. Transformation in yeast: development of a hybrid cloning vector and isolation of the *CAN1* gene. Gene 8:121–133.
8. Crous, J. M., I. S. Pretorius, and W. H. van Zyl. 1995. Cloning and expression of an *Aspergillus kawachii* endo-1,4-β-xylanase gene in *Saccharomyces cerevisiae*. Curr. Genet. 28:467–473.
9. Dekker, R. F. H. 1983. Bioconversion of hemicellulose: aspects of hemicellulase production by *Trichoderma reesei* QM 9414 and enzymic saccharification of hemicellulose. Biotechnol. Bioeng. 25:1127–1146.
10. Farkas, V., M. Liskovi, and P. Biely. 1985. Novel media for detection of microbial producers of cellulase and xylanase. FEMS Microbiol. Lett. 28: 137–140.
11. Gomori, G. 1959. Preparation of buffers for use in enzyme studies. Methods Enzymol. 1:138–146.
12. Higuchi, R. 1990. Recombinant PCR, p. 177–183. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols: a guide to methods and applications. Academic Press, Inc., San Diego, Calif.
13. Hill, J., K. A. Ian, G. Donald, and D. E. Griffiths. 1991. DMSO-enhanced whole cell yeast transformation. Nucleic Acids Res. 19:5791.
14. Hill, J. E., A. M. Myers, T. J. Koerner, and A. Tzagoloff. 1986. Yeast/*E. coli* shuttle vectors with multiple unique restriction sites. Yeast 2:163–167.
15. Hitzeman, R. A., F. E. Hagie, J. S. Hayflick, C. Y. Chen, P. H. Seeburg, and R. Derynck. 1982. The primary structure of the *Saccharomyces cerevisiae* gene for 3-phosphoglycerate kinase. Nucleic Acids Res. 10:7791–7808.
16. Hoffman, C. S., and F. Winston. 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene 57:267–272.
17. Innis, M. A., and D. H. Gelfand. 1990. Optimization of PCRs, p. 3–12. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White (ed.), PCR protocols: a guide to methods and applications. Academic Press, Inc., San Diego, Calif.
18. Jeffries, T. W. 1983. Utilization of xylose by bacteria, yeasts, and fungi. Adv. Biochem. Eng. 27:1–32.
19. Kern, L., J. De Montigny, R. Jund, and F. Lacroute. 1990. The *FUR1* gene of *Saccharomyces cerevisiae*: cloning, structure and expression of wild-type and mutant alleles. Gene 88:149–157.
20. Kozak, M. 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem. 266:19867–19870.
21. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680–685.
22. Laing, E., and I. S. Pretorius. 1992. Synthesis and secretion of an *Erwinia chrysanthemi* pectate lyase in *Saccharomyces cerevisiae* regulated by different combinations of bacterial and yeast promoter and signal sequences. Gene 121:35–45.
23. Maat, J., M. Roza, J. Verbakel, H. Stam, M. J. Santos da Silva, M. Bosse, M. R. Egmond, M. L. D. Hagemans, R. F. M. V. Gorcom, J. G. M. Hessing, C. A. M. J. J. van der Hondel, and C. V. Rotterdam. 1992. Xylanases and their application in bakery, p. 349–360. *In* J. Visser, G. Beldman, M. A. Kusters-van Someren, and A. G. J. Voragen (ed.), Xylans and xylanases. Elsevier Science, Amsterdam.
24. Mandels, M., J. Weber, and R. Parizek. 1971. Enhanced cellulase production by a mutant of *Trichoderma viride*. Appl. Microbiol. 21:152–154.
25. Matsuo, M., and T. Yasui. 1984. Purification and some properties of β-xy-

losidase from *Trichoderma viride*. Agric. Biol. Chem. 48:1845–1852.
26. **Miller, G. L., R. Blum, W. E. Glennon, and A. L. Burton.** 1960. Measurement of carboxymethylcellulase activity. Anal. Biochem. 2:127–132.
27. **Nevalainen, H., A. Harkki, M. Penttilä, M. Salohelmo, T. T. Teeri, and J. K. C. Knowles.** 1990. *Trichoderma reesei* as a production organism for enzymes for the pulp and paper industry. p. 593–599. *In* T. K. Kirk and H. M. Chang (ed.), Biotechnology in pulp and paper manufacture. Butterworth-Heinemann, Boston.
28. **Penttilä, M., L. André, P. Lehtovaara, M. Bailey, T. T. Teeri, and J. K. C. Knowles.** 1988. Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae*. Gene 63:103–112.
29. **Penttilä, M., T. T. Teeri, H. Nevalainen, and J. K. C. Knowles.** 1990. The molecular biology of *Trichoderma reesei* and its application to biotechnology. p. 85–102. *In* J. F. Peberdy, C. E. Caten, J. E. Ogden, and J. W. Bennett (ed.), Applied molecular genetics of fungi. Cambridge University Press, Cambridge.
30. **Price, V. L., W. E. Taylor, W. Clevenger, M. Worthington, and E. T. Young.** 1990. Expression of heterologous proteins in *Saccharomyces cerevisiae* using the *ADH2* promoter. Methods Enzymol. 185:308–318.
31. **Puls, J., and J. Schuseil.** 1993. Chemistry of hemicelluloses: relationship between hemicellulose structure and enzymes required for hydrolysis. p. 1–27. *In* M. P. Coughlan and G. P. Hazlewood (ed.), Hemicellulose and hemicellulases. Portland Press, London.
32. **Romanos, M. A., C. A. Scorer, and J. J. Clare.** 1992. Foreign gene expression in yeast: a review. Yeast 8:423–488.
33. **Rose, M. D., F. Winston, and P. Hieter.** 1990. Methods in yeast genetics: a laboratory course manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **Rothstein, R. J.** 1983. One step gene disruption in yeast. Methods Enzymol. 101:202–211.
35. **Russell, D. W., M. Smith, V. M. Williamson, and E. T. Young.** 1983. Nucleotide sequence of the yeast alcohol dehydrogenase II gene. J. Biol. Chem. 258:2674–2682.
36. **Saarelainen, R., M. Paloheimo, R. Fagerström, P. L. Suominen, and K. M. H. Nevalainen.** 1993. Cloning, sequencing and enhanced expression of the *Trichoderma reesei* endoxylanase II (pI 9) gene *xln2*. Mol. Gen. Genet. 241:497–503.
37. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491.
38. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
40. **Sheir-Neiss, G., and B. S. Montenecourt.** 1984. Characterization of the secreted cellulases of *Trichoderma reesei* wild type and mutants during controlled fermentations. Appl. Microbiol. Biotechnol. 20:46–53.
41. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.
42. **Tenkanen, M., J. Buchert, J. Puls, K. Poutanen, and L. Viikari.** 1992. Two main xylanases of *Trichoderma reesei* and their use in pulp processing. p. 547–550. *In* J. Visser, G. Beldman, M. A. Kusters-van Someren, and A. G. J. Voragen (ed.), Xylans and xylanases. Elsevier Science, Amsterdam.
43. **Törrönen, A., A. Harkki, and J. Rouvinen.** 1994. Three-dimensional structure of endo-1,4-β-xylanase II from *Trichoderma reesei*: two conformational states in the active site. EMBO J. 13:2493–2501.
44. **Törrönen, A., R. L. Mach, R. Messner, R. Gonzalez, N. Kalkkinen, A. Harkki, and C. P. Kubicek.** 1992. The two major xylanases from *Trichoderma reesei*: characterization of both enzymes and genes. Bio/Technology 10:1461–1465.
45. **Van Paridon, P. A., J. C. P. Boonman, G. C. M. Selten, C. Geerse, D. Barug, P. H. M. De Bot, and G. Hemke.** 1992. The application of fungal endoxylanase in poultry diets. p. 371–378. *In* J. Visser, G. Beldman, M. A. Kusters-van Someren, and A. G. J. Voragen (ed.), Xylans and xylanases. Elsevier Science, Amsterdam.
46. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268.
47. **Wong, K. K. Y., and J. N. Saddler.** 1992. *Trichoderma* xylanases, their properties and application. p. 171–186. *In* J. Visser, G. Beldman, M. A. Kusters-van Someren, and A. G. J. Voragen (ed.), Xylans and xylanases. Elsevier Science, Amsterdam.
48. **Wong, K. K. Y., and J. N. Saddler.** 1992. *Trichoderma* xylanases, their properties and application. Crit. Rev. Biotechnol. 12:413–435.
49. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119.

Directed evolution to produce an alkalophilic variant from a *Neocallimastix patriciarum* xylanase

Yew-Loom Chen, Tsung-Yin Tang, and Kuo-Joan Cheng

Abstract: The catalytic domain of a xylanase from the anaerobic fungus *Neocallimastix patriciarum* was made more alkalophilic through directed evolution using error-prone PCR. Transformants expressing the alkalophilic variant xylanases produced larger clear zones when overlaid with high pH, xylan-containing agar. Eight amino acid substitutions were identified in six selected mutant xylanases. Whereas the wild-type xylanase exhibited no activity at pH 8.5, the relative and specific activities of the six mutants were higher at pH 8.5 than at pH 6.0. Seven of the eight amino acid substitutions were assembled in one enzyme (xyn-CDBFV) by site-directed mutagenesis. Some or all of the seven mutations exerted positive and possibly synergistic effects on the alkalophilicity of the enzyme. The resulting composite mutant xylanase retained a greater proportion of its activity than did the wild type at pH above 7.0, maintaining 25% of its activity at pH 9.0, and its retention of activity at acid pH was no lower than that of the wild type. The composite xylanase (xyn-CDBFV) had a relatively high specific activity of 10 128 μmol glucose·min^{-1}·(mg protein)$^{-1}$ at pH 6.0. It was more thermostable at 60°C and alkaline tolerant at pH 10.0 than the wild-type xylanase. These properties suggest that the composite mutant xylanase is a promising and suitable candidate for paper pulp bio-bleaching.

Key words: xylanase, *Neocallimastix patriciarum*, alkalophilicity, random mutagenesis, directed evolution.

Résumé : Le domaine catalytique d'une xylanase du champignon anaérobe *Neocallimastix patriciarum* a été rendu plus alcalophilique par évolution dirigée en utilisant un PCR à retranscription imparfaite. Les transformants exprimant des variantes de xylanases alcalophiliques ont produit des zones claires plus étendues lorsqu'elles étaient recouvertes avec un agar-xylane à pH élevé. Nous avons identifié huit substitutions d'acides aminés dans six xylanases mutantes. Alors que la xylanase de type sauvage ne manifestait aucune activité à un pH de 8,5, les activités relatives et spécifiques des six mutants étaient plus élevées à pH 8,5 qu'à pH 6,0. Sept des huit substitutions d'acides aminés ont été assemblées dans une enzyme (xyn-CDBFV) par mutagenèse dirigée. Certaines ou toutes les sept mutations ont eu des effets positifs et vraisemblablement synergiques sur l'alcalophilicité de l'enzyme. Le mutant composite de la xylanase qui en est découlé conservait une plus grande proportion de son activité que l'enzyme de type sauvage à un pH supérieur à 7,0; elle a maintenu 25 % de son activité à un pH de 9,0, et la conservation de son activité à des pH acides n'était pas inférieure à celle du type sauvage. La xylanase composite (xyn-CDBFV) avait une activité spécifique relativement élevée de 10 128 μmol glucose·min^{-1}·(mg protéine)$^{-1}$ à un pH de 6,0. Elle était plus thermostable à 60°C et plus tolérante à l'alcalinité à pH 10,0 que la xylanase de type sauvage. Ces propriétés indiquent que la xylanase composite représente un candidat prometteur adapté au bio-blanchiment de la pâte à papier.

Mots clés : xylanase, *Neocallimastix patriciarum*, alcalophilicité, mutagénèse aléatoire, évolution dirigée.

[Traduit par la Rédaction]

Introduction

Endo-β-1,4-xylanase (EC 3.2.1.8) catalyzes the hydrolysis of xylan, a plant cell wall heteropolysaccharide that contains a backbone of 1,4-linked xylose residues (Kulkarni et al. 1999). Xylanases have been used industrially to bio-bleach paper pulp (Buchert et al. 1994; Bajpai 1999), as an animal feed supplement (Uhlig 1998), to clarify fruit juice (Biely 1985), and to enhance the quality of baked products (Maat et al. 1992). For environmental reasons, xylanases are desirable in that they reduce the amount of chlorine and chlorine dioxide used for bleaching paper pulp. During the bleaching of kraft paper pulp, the lignin in wood chips is removed by sequential treatments with chlorine, chlorine dioxide, and NaOH (Viikari et al. 1994). The chlorine and chlorine dioxide create persistent organic chemicals that are toxic to organisms in the waterways close to paper plants and may present health risks to humans as well. Pretreating paper pulp with xylanases can enhance the efficiency of the chemical extraction of lignin (Viikari et al. 1994; Bim and Franco 2000; Georis et al. 2000) and so reduce the amounts of chlorine and chlorine dioxide

Received June 4, 2001. Accepted October 5, 2001. Published on the NRC Research Press Web site at http://cjm.nrc.ca on December 3, 2001.

Y.-L. Chen,[1] T.-Y. Tang, and K.-J. Cheng.[2] Institute of BioAgricultural Sciences, Academia Sinica, No. 128, Sec. 2, Academia Sinica Road, Nankang 11529, Taipei, Taiwan, Republic of China.

[1]Present address: Graduate Institute of Life Science, National Defense Medical Center, Taipei, Taiwan, Republic of China.
[2]Corresponding author (e-mail: yewloom@hotmail.com).

 DOI: 10.1139/cjm-47-12-1088

Fig. 1. Amino acid sequence of xyn-CD/WT xylanase expressed in pGEX-4T1 vector. Every fifth amino acid (numbers immediately above) is depicted in bold type. The eight amino acids depicted in bold, italicized type (numbers 24, 97, 128, 164, 168, 208, 214, and 218) are those substituted by others in the alkalophilic mutants listed in Table 1.

```
1    5    10   15   20   25   30
G S P E F M L A Q S F C S S A S H S G Q S V K E T G N K V G

31   35   40   45   50   55   60
T I G G V G Y E L W A D S G N N S A T F Y S D G S F S C T F

61   65   70   75   80   85   90
Q N A G D Y L C R S G L S P D S T K T P S Q I D R M K A D F

91   95   100  105  110  115  120
K L V K Q N I S N V G Y S Y V G V Y G W T R S P L V E Y Y I

121  125  130  135  140  145  150
V D N W L S P S P P G D W V G N K K H G S F T I D G A Q Y T

151  155  160  165  170  175  180
V Y E N T R T G P S I D G N T T F K Q Y F S I R Q Q A R D C

181  185  190  195  200  205  210
G T I D I S A H F D Q W E K L G M T M G K L H E A K V L G E

211  215  220  225  230  235  240
A G N G N G G V S G T A D F P Y A K V Y I G D L E R P H R D
```

required. Because pulp bleaching entails high temperatures and high pH, the optimum xylanase for use in this application would be thermophilic, thermostable, alkalophilic, and stable in alkaline environments (Shoham et al. 1992; Bim and Franco 2000).

The potential usefulness of these enzymes in industry has spurred considerable research efforts toward producing more thermostable, thermophilic, and alkalophilic xylanases. Two approaches have been taken: screening organisms from various sources in nature for high xylanase activity, and engineering known xylanases to improve their characteristics. Possibly, both approaches are necessary to obtain ideal xylanases. Thermophilic xylanases have been isolated from the hyperthermophilic bacterium *Thermotoga maritima* (Simpson et al. 1991), from *Bacillus* spp. (Harris et al. 1997), and from the thermophilic fungus *Thermomyces lanuginosus* (Gruber et al. 1998). These xylanases are insufficiently alkalophilic for bio-bleaching, however. Their pH optima are between 5.5 and 6.0 compared with pH 9.0, which is the ideal (Bim and Franco 2000).

Random mutagenesis (error-prone PCR) has been used successfully to improve specific properties of enzymes, particularly thermostability. Giver et al. (1998) made a thermostable esterase using this technique together with recombination. In the same way, Zhao and Arnold (1999) made a thermostable, thermophilic thermitase. Error-prone PCR has been used to increase the thermostability (and specific activity) of a phospholipase A1 (Song and Rhee 2000) and a cholesterol oxidase from *Streptomyces* (Nishiya et al. 1997). The thermostability of a xylanase was increased through chemical (random) mutagenesis (Arase et al. 1993). Research into improving alkalophilicity of enzymes has been more limited, however. Russell and Fersht (1987) shifted the pH-activity profile of subtilisin by site-directed mutagenesis based on rational theoretical design. The pH optimum of a D-xylose isomerase was shifted downward 0.5 to 0.8 units by similar rational design (Cha and Batt 1998).

A xylanase from the ruminal fungus *Neocallimastix patriciarum* was characterized (Tamblyn Lee et al.1993), and the gene encoding it was subcloned, sequenced, and subjected to structural analysis (Liu et al. 1999). This study was conducted to explore the possibility of using random mutagenesis and site-directed mutagenesis to shift the pH optimum of the *N. patriciarum* xylanase from acid (pH 6.2) to alkaline pH and (or) to increase its specific activity at alkaline pH to render the enzyme better suited to the harsh conditions of paper pulp bio-bleaching.

Materials and methods

Construction of plasmid pxyn-CD/WT

The catalytic domain of the *xynC* xylanase gene from *N. patriciarum* (Liu et al. 1999) was amplified for expression using the plasmid pGEXxynC as a template and the oligonucleotide primers 5′-xyn1 (5′-GAGAGAATTCATGCTTGCTCAAAGTTTC-TGT-3′) and 3′-xyn2 (5′-CGCGCTCGAGATCACCAATGTAAA-CCTT-3′). The PCR product was digested with *Eco*RI and *Xho*I and cloned as a translational fusion into *Eco*RI-, *Xho*I-digested pGEX-4T-1 plasmid (Pharmacia Biotech Inc., Piscataway, N.J.). The resultant expression vector was named pxyn-CD/WT. When expressed in *Escherichia coli*, the gene product of the catalytic domain of *xynC*, designated as xyn-CD/WT, became a fusion protein with glutathione *S*-transferase. Thrombin cleavage allowed release of the recombinant xyn-CD/WT protein, which has eight extra amino acid residues (GSPEFMLA) at the N-terminal and seven residues (LERPHRD) at the C-terminal (Fig. 1).

Error-prone PCR and construction of the libraries of variant xylanase genes

To generate a library of variants of the gene encoding xyn-CD/WT, we introduced mutations using error-prone PCR. The reaction mixture comprised 50–100 ng pxyn-CD/WT, 50 pmol forward primer (xyn1), 50 pmol reverse primer (xyn2), and 2 U of *Taq* DNA polymerase (Promega, Madison, Wis.) in 100 µL of 1× reaction buffer. It also contained 1.5 mM $MgCl_2$, 0.2 mM dCTP, 0.2 mM dTTP, 0.04 mM dATP, and 0.04 mM dGTP. Thermocycler settings (TouchDown™ Temperature Cycling System, Hybaid, Middlesex, U.K.) were 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s (50 cycles). The PCR products (~700 bp) were isolated from a 1% agarose gel and purified, using a QIAEX II gel extraction kit (Qiagen, Valencia, Calif.), and were then cloned and ligated with T4 DNA ligase (Promega) into *Eco*RI-, *Xho*I-digested pGEX-4T-1 vector. The resulting plasmids were transferred into *E. coli* BL21-DE3 (Novagen, Madison, Wis.) at 42°C heat shock transformation.

Screening for alkalophilic xylanase variants

Transformed cells derived from the library of variants of xyn-CD/WT were plated onto Luria-Bertani (LB) agar plates (Miller 1972) containing ampicillin (100 µg/mL) and were incubated at 37°C for approximately 12 h. Colonies were overlaid with molten 1% agar (50°C) in 0.1 M disodium tetraborate buffer containing 0.1% oat spelts xylan (Sigma Chemical Co., St. Louis, Mo.). The pH of the overlay buffer was adjusted to 9.0 during the first selection cycle and to 10.0 during the second selection cycle. After overnight incubation at room temperature, the plates were stained with 1% (w/v) Congo red (Sigma) for 1 h and destained with 1 M NaCl solution for 2 h (Teather and Wood 1982). The colonies with relatively larger halos (i.e., zones of xylan degradation) were selected for subculturing; they were the strains containing plasmids expressing xylanase variants with increased alkalophilicity. The selected clones were DNA sequenced through a *Taq* DyeDeoxy™

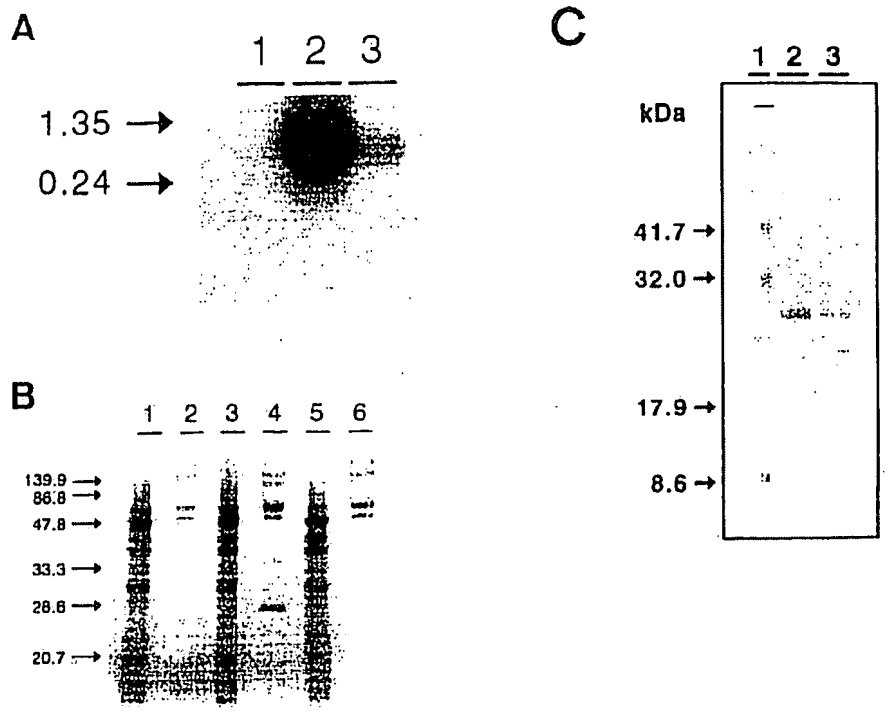


FIG. 5. (A) Northern blot analysis of total RNA isolated from *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG1) (lane 1), *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5) (lane 2), and *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6) (lane 3). Molecular sizes are indicated in kilobases. The 0.7-kb *XYN2* DNA fragment was used as a α-^{32}P-labelled probe. (B) SDS-PAGE of the Xyn2 protein produced by *S. cerevisiae* Y294 strains. Lanes: 1 and 2, *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG1); 3 and 4, *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5); 5 and 6, *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6). Lanes 1, 3, and 5 contain the intracellular protein fractions, and lanes 2, 4, and 6 contain the extracellular fractions. The band corresponding to the β-xylanase enzyme is at ca. 27 kDa. (C) SDS-PAGE of endoglycosidase F-treated Xyn2 protein produced by *S. cerevisiae* Y294. Lanes: 1, protein molecular markers; 2, untreated extracellular fraction of *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5); 3, endoglycosidase F-treated extracellular fraction of *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5).

mRNA produced by *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5) probably leads to higher levels of enzyme compared with those in *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6) (Fig. 5B). Treatment of the 27-kDa protein species with endoglycosidase F generated a new protein species of 21 kDa that corresponds to the molecular mass of native Xyn2 xylanase isolated from *T. reesei* (44) (Fig. 5C).

β-Xylannse activity. The β-xylanase-producing yeast strains and *T. reesei* QM 6a were analyzed for their ability to secrete biologically active β-xylanases over 125- and 250-h periods, respectively (Fig. 6). For optimal production of β-xylanase, *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6) was cultured on YPD medium with 0.8% glucose (15), *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5) was cultured on YPD medium with 0.8% galactose, and *T. reesei* was cultured on basal medium containing 0.3% oat spelts xylan (25). Lower concentrations of glucose and galactose were used to reduce the reducing-sugar background at the start of the growth curve, enabling the assessment of β-xylanase activity in the culture supernatants. The highest β-xylanase activity for *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6) and *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5) was recorded after ca. 70 and ca. 80 h, respectively (Fig. 6A and B). *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6) showed a peak β-xylanase activity of 160 nkat/ml, and *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG5) showed a activity of 1,200 nkat/ml. β-Xylanase activity was detected earlier in *S. cerevisiae* Y294(*fur1*::*LEU2* pDLG6), as a result of the constitutive *PGK1* promoter. It also reached a higher cell density in a shorter period than in *S. cerevisiae* Y294 (*fur1*::*LEU2* pDLG5), probably because glucose is preferred to galactose as a carbon source for *S. cerevisiae* growth. In *T. reesei* QM 6a, β-xylanase activity was detected after about 65 h, with the highest levels measured at ca. 450 nkat/ml after about 180 h (Fig. 6C).

Effects of pH and temperature on β-xylanase activity. The recombinant β-xylanase activity peaked between pH 4 and 6, with the highest activity measured at pH 6 in 50 mM citrate buffer (Fig. 7A). The optimum temperature for this enzyme was at 60°C (Fig. 7B). Although the highest β-xylanase activity was measured at 60°C, the enzyme is not stable at this temperature (Fig. 7C). The β-xylanase activity decreased by more than 50% after a 5-min incubation at 60°C; less than 10% activity could be measured after a 10-min incubation (Fig. 7D). However, the recombinant enzyme is relatively stable at 50°C: more than 90% activity remained after 30 min at this temperature (Fig. 7C).

DISCUSSION

mRNA was isolated from the xylanolytic fungus *T. reesei*, and the *XYN2* gene encoding the main β-xylanase, Xyn2, was amplified with the aid of sequence-specific PCR primers. The DNA sequence was verified and compared with the DNA sequence published by Törrönen et al. (44) and with the DNA sequences available in the GenBank database (release 87.0, 13

Table 1. Designations, amino acid substitutions, and specific activities of the wild-type and mutant xylanases from *Neocallimastix patriciarum*.

Xylanase	Source	Amino acid substitutions	Specific activity* at 47°C		
			pH 6.0†	pH 8.5†	% activity retained
xyn-CD/WT	Wild-type xylanase	None	4 343	0	0
			(98.5)	–	–
xyn-O9D	First round of PCR	I97S	3 460	47	1.36
			(83.5)	(7.3)	
xyn-C2	Second round of PCR	I97S, K168N, G214V	2 581	97	3.06
			(65.8)	(45.2)	
xyn-D8	Second round of PCR	I97S, N164D	4 110	252	6.13
			(102.0)	(15.9)	
xyn-D2	Second round of PCR	I97S, S128F, L208F	2 831	306	10.81
			(54.5)	(35.8)	
xyn-B5	Second round of PCR	I97S, E24V	3 710	404	10.89
			(164.2)	(26.4)	
xyn-C3	Second round of PCR	I97S, V218A	3 934	534	3.57
			(123.4)	(42.9)	
xyn-CDBFV	Site-directed mutagenesis	I97S, K168N, G214V, N164D, S128F, E24V, V218A	10 128	4542	44.85
			(302.8)	(315.2)	

*Specific activity is expressed as micromoles of glucose released from oat spelts xylan per minute per milligram of protein under standard assay conditions (10 min; 1% (w/v) substrate). Values shown are the means of quadruplicate determinations, except xyn-C3 (pH 6.0) and xyn-C2 (pH 8.5), which were determined in triplicate, and xyn-O9D (pH 8.5), which was determined in duplicate.
†Values in parentheses are standard deviations of the means immediately above.

Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, Calif.). The gene products (xylanases) from those variants possessing amino acid mutation(s) were purified, as described below, for the pH–activity profile assay.

Two step selection process

From the first mutational library constructed and screened at pH 9.0, one clone (*E. coli* xyn-O9D) was selected based on the size of its xylan-clearing zone (halo). The xylanase from this clone was purified, as described below, and its xylanolytic activity was assessed at pH from 4.0 to 9.0. The pH activity profile confirmed that the alkalophilicity of the enzyme was increased over that of the wild type. The xylanase-encoding gene from this mutant was then used as template for a second round of error-prone PCR, producing a second mutational library. Screening of this second library for xylanase activity at pH 10.0 yielded an additional five mutants even more alkalophilic than *E. coli* xyn-O9D.

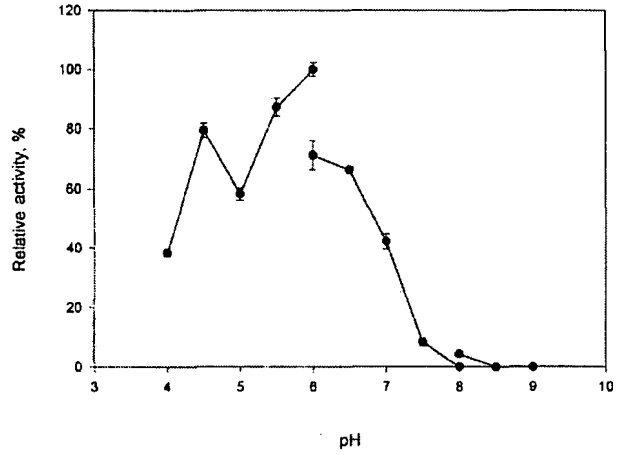
Site-directed mutagenesis

DNA sequencing of the six mutant xylanase genes (xyn-O9D, xyn-C2, xyn-D8, xyn-D2, xyn-B5, and xyn-C3; Table 1) revealed amino acid substitutions at eight sites in the genes (Fig. 1). Using site-directed mutagenesis, these eight mutations were assembled into a series of composite mutant xylanases, which had various combinations of the eight mutations. Site-directed mutageneses were performed, using the ExSite™ PCR-based site-directed mutagenesis kit (Stratagene Cloning Systems, Inc., La Jolla, Calif.). Complementary primers for PCR containing the desired DNA changes were constructed and included in reaction mixtures in which 40 µL distilled H_2O, 5 µL 10× pfu-turbo polymerase buffer, 1 µL forward primer (125 ng/µL), 1 µL reverse primer (125 ng/L), 1 µL dNTPs (10 mM), 1 µL plasmid (10–50 ng), and 1 µL pfu-turbo polymerase (2.5 U/µL) were combined in a total reaction volume of 50 µL. Thermocycler settings (TouchDown™, Hybaid) were as follows: 95°C for 4 min, then 15 cycles at 95°C for 30 s, 58°C for 30 s, and 68°C for 13 min. The final primer extension was conducted at 70°C for 10 min. *Dpn*I (1 µL) was added to the reaction, followed by a further incubation at 37°C for 2 h. The resulting plasmids directly from the final reaction mixture were transformed into competent cells of *E. coli* DH5α, and the resultant colonies were DNA sequenced to identify the clones containing the desired substitutions.

Protein purification and pH-activity profiles

Escherichia coli BL21-DE3 was used for overexpression of the xylanases. Cells containing the plasmid of interest were cultured overnight in LB broth containing ampicillin (100 µg/mL), then diluted 1:100 in fresh broth to generate 100 mL of culture and incubated for a further 2 h. The cells were induced by adding 100 µL of 0.1 M stock solution of isopropyl-β-D-thiogalactopyranoside (0.1 mM) and further incubated at 37°C for 2–3 h. Following induction, the cultures were centrifuged (7700 × *g*; 10 min; 4°C), and the pellets were suspended in 5 mL of single-strength phosphate buffered saline (1× PBS) for lysis by sonication (20 pulses of 10 s). The lysates were centrifuged (12 000 × *g*; 10 min; 4°C) and 5-mL aliquots were combined with 200 µL of glutathione sepharose 4B slurry (Pharmacia) for 2 h to bind the GST-fusion proteins. The resultant beads were washed four or five times with PBS, which excluded almost all nonspecific binding as determined by SDS–PAGE assay. The target GST-fused xylanases were released from the beads by incubation for 7–12 h at room temperature with 315 µL of PBS buffer, which contained 15 U of thrombin protease (Pharmacia). Protein concentrations in the resulting samples were determined, using a Bio-Rad protein assay kit (Bio-Rad Laboratories, Richmond, Calif.), against a bovine serum albumin standard. Activities of the xylanases were determined at pH ranging from 4.0 to 9.0 under standardized assay conditions using the following three buffers: 0.05 M citric acid, for pH 4.0, 4.5, 5.0,

Fig. 2. Activity profile of xyn-CD/WT xylanase at pH 4.0–9.0, determined in three 0.05 M buffers (citric acid, for pH 4.0–6.0; sodium phosphate, for pH 6.0–8.0; and Tris-HCl, for pH 8.0–9.0). Activity was expressed as micromoles of glucose equivalents released from xylan per milligram protein per minute under standard assay conditions. The highest activity, observed in pH 6.0 citric acid buffer, was considered as 100% activity, and others are plotted here as percentages of that activity. Values shown are the means of quadruplicate determinations. Bars indicate standard deviations. Where not visible, bars fall within symbols.



5.5, and 6.0; 0.05 M sodium phosphate, for pH 6.0, 6.5, 7.0, 7.5, and 8.0; and 0.05 M Tris-HCl for pH 8.0, 8.5, and 9.0. Substrate (oat spelts xylan, X-0627, Sigma) was included in each buffer at 1% (w/v). Incubations (100 µL total volume) were conducted at 47°C for 10 min, then were terminated by adding 50 µL of stopping solution (a combined aqueous solution of 25% (w/v) K_2CO_3 and 5% (w/v) $Na_2S_2O_3$), followed by 50 µL of 0.3% (w/v) 3,6-dinitrophthalic acid. The resultant mixtures were boiled for 10 min, stored on ice for at least 5 min, and their optical densities at 450 nm were measured with a MultiSkan microtitre plate reader (Labsystem, Helsinki, Finland). The OD_{450} values were converted to micromoles of glucose equivalents, using a standard curve of glucose versus OD_{450}. Amounts of xylanase used ranged from 0.0035 to 0.008 µg protein in the 100-µL reaction mixture, depending on the specific activity of each xylanase.

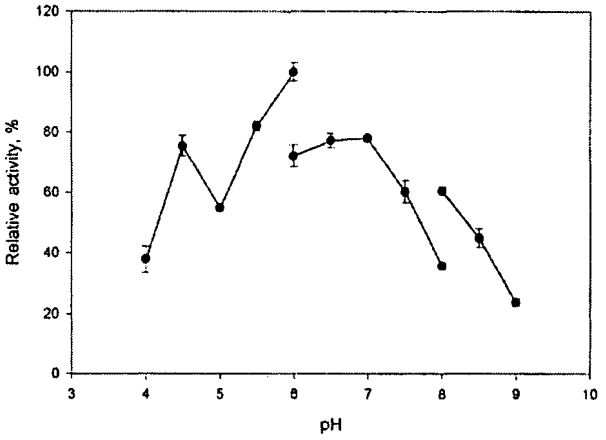
Assay of resistance of xyn-CDBFV xylanase to an alkaline environment

The xyn-CDBFV xylanases were incubated at room temperature in 0.1 M 2-amino-2-methyl-1-propanol at pH 10.0 (0.0324 µg protein in 50 µL buffer) for 0, 2, 4, 6, 12, and 24 h, then 50 µL of 2% (w/v) xylan solution was added. Incubation was continued at 47°C for 10 min, then xylanase activity (measured as release of glucose equivalents) was determined, as described above. Activities in the 0-h samples were considered as 100%, and activities at the other incubation time points were expressed as percentages of this control.

Assay of thermostability and temperature optima of the xylanases

Thermostabilities of xyn-CD/WT and xyn-CDBFV xylanases were determined by incubating the enzymes in 50 µL of 0.1 M citric acid buffer (pH 6.0) at 55 or 60°C for 10, 20, 40, 60, 80, 100, or 120 min, chilling on ice for at least 20 min, and then determining xylanase activity, using the same enzyme concentrations as

Fig. 3. Activity profile of xyn-CDBFV xylanase at pH 4.0–9.0, determined as described for xyn-CD/WT in Fig. 2. The highest activity (in citric acid buffer, pH 6.0) was considered as 100% activity, and others are plotted as percentages of that activity. Values shown are the means of quadruplicate determinations. Bars indicate standard deviations. Where not visible, bars fall within symbols.



were used for pH optima. The cooled samples were equilibrated to room temperature for 20 min prior to xylanase assays. Incubation conditions were those described previously: 10 min at 47°C in 0.05 M citric acid buffer (pH 6.0) containing 1% (w/v) xylan. The temperature optima of the two xylanases were determined by incubating them in 0.05 M citric acid buffer (pH 6.0) containing 1% xylan for 10 min at 42, 47, 52, 57, 62, 67, 72, and 77°C. Reactions were terminated and activities measured, as described above.

Results and discussion

Evolution of a more alkalophilic xylanase, xyn-CDBFV, by random mutagenesis and site-directed mutagenesis

Error-prone PCR conducted on the xyn-CD/WT xylanase gene cloned into pGEX-4T-1 produced a mutation library from which mutant forms of the wild-type *N. patriciarum* xylanase with enhanced alkalophilicity were effectively identified by agar overlay screening. The single clone selected from the first mutation library, xyn-O9D, displayed enhanced alkalophilicity relative to the parent peptide, xyn-CD/WT (Fig. 2 and Table 1). At pH 8.5 and 47°C, the specific activity of xyn-O9D was 47 µmol glucose·min^{-1}·(mg protein)$^{-1}$, whereas xyn-CD/WT exhibited no activity at that pH. Agar overlay screening (at pH 10.0) of a second mutation library constructed from the xyn-O9D gene yielded five additional mutants more alkalophilic than xyn-O9D (xyn-C2, xyn-D8 xyn-D2, xyn-B5, and xyn-C3; Table 1). Site-directed mutagenesis was effective for compiling the eight amino acid substitutions identified by DNA sequencing of the six mutants into a series of composite mutant xylanases. One of the composite mutant xylanases, xyn-CDBFV (with seven amino acid mutations), proved to be substantially more alkalophilic than any of the six mutants produced by error-prone PCR (Table 1).

Fig. 4. Temperature profiles of xyn-CD/WT (●) and xyn-CDBFV (■) xylanases, determined in 10-min incubations at pH 6.0 (citric acid buffer). For each xylanase, activities are shown as percentages of the maximum activities, recorded at 47°C for xyn-CD/WT and at 62°C for xyn-CDBFV. Values shown are means of triplicate determinations.

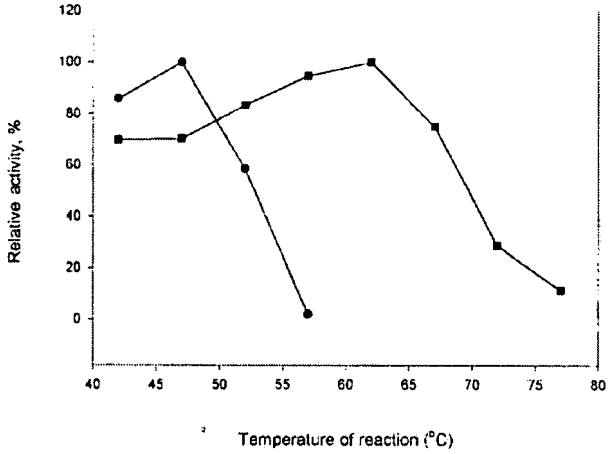


Fig. 5. Thermostability of xyn-CD/WT (●) and xyn-CDBFV (■) at 60 EC, determined at pH 6.0 (citric acid buffer). Activities are plotted as percentages of the maximal activity, which was observed in the 10-min incubation, for both enzymes. Values shown are the means of quadruplicate determinations. Bars indicate standard deviations. Where not visible, bars fall within symbols.

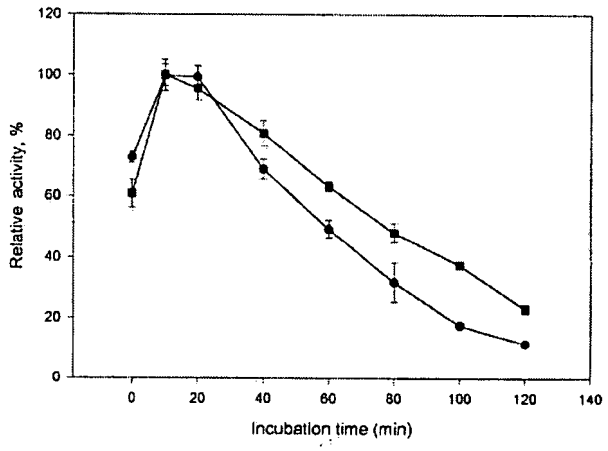
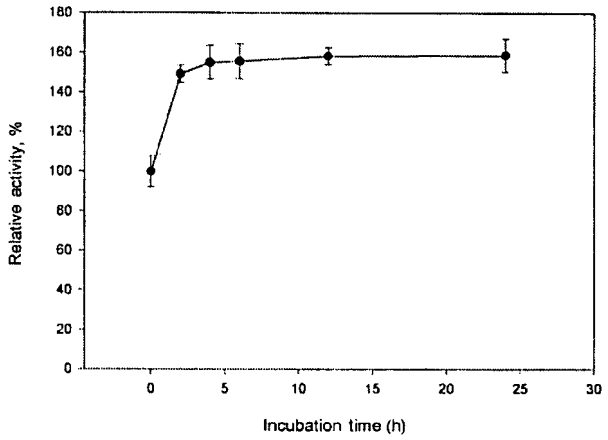


Fig. 6. Alkaline resistance of xyn-CDBFV xylanase at pH 10 in 0.1 M 2-amino-2-methyl-1-propanol (AMP) buffer. The xyn-CDBFV protein samples were in pH 10, 0.1 M AMP buffer and incubated for 0, 2, 4, 6, 12, 24 h at room temperature. Four aliquots of protein samples were assayed for each time course. The activity produced by 0 h alkaline treatment samples is regarded as the 100% control. The average specific activity of 0-h samples was about 67 μmol glucose·min^{-1}·(mg protein)$^{-1}$.



pH profiles of the eight mutant xylanases

The eight purified xylanase proteins, xyn-CD/WT, xyn-O9D, xyn-B5, xyn-C2,xyn-D2, xyn-D8, xyn-C3, and xyn-CDBFV (Table 1) were assayed for xylanase activity at pH between 4.0 and 9.0. The pH optimum of the wild-type xylanase, xyn-CD/WT, was 6.0 (Fig. 2). Its activity was reduced to only about 10% of the optimum at pH 7.5 and was inhibited entirely at pH 8.5 or 9.0. The highest activity of composite xylanase, xyn-CDBFV, was also observed at pH 6.0 (Fig. 3), but in contrast to xyn-CD/WT, this mutant retained about 60% of its highest activity at pH 7.5 and maintained about 45% and 25% of its activity at pH 8.5 and 9.0, respectively. It can therefore be concluded that xyn-CDBFV is substantially more alkalophilic than is xyn-CD/WT. Specific activities of the six PCR-generated mutants at pH 8.5 and their relative retentions of activity at pH 8.5 versus pH 6.0 were intermediate to the wild type (xyn-CD/WT) and the composite mutant (xyn-CDBFV) xylanases (Table 1). The specific activity of xyn-CDBFV at pH 8.5 (4542 μmol glucose·min^{-1}·(mg protein)$^{-1}$) was approximately 8.5 times higher than that of the best of the six PCR mutants, whose specific activities ranged from 47 to 534 μmol glucose·min^{-1}·(mg protein)$^{-1}$. Substitution I97S, which was common to all the mutants, conferred retention of a small amount of activity at pH 8.5 (xyn-O9D versus xyn-CD/WT). Except for mutant xyn-C2, combining substitution I97S with one or two others resulted in specific activities at pH 8.5 that were 5.4–11.4 times higher than that in the single substitution mutant (xyn-O9D). Further, compilation of seven amino acid substitutions in a composite mutant (xyn-CDBFV) increased specific activity at pH 8.5 by over 96 times that exhibited by the single substitution mutant. It may be that some or all of the seven amino acid substituants in xyn-CDBFV exert synergistic positive effects on the alkalophilicity of the xylanase. One of the substitutions, L208F in the xyn-D2 mutant, appeared to have a negative effect on alkalophilicity. A composite mutant containing all eight substitutions had lower alkalophilicity and lower specific activity at pH 8.5 than did composite mutant xyn-CDBFV, which did not contain L208F (data not shown).

Three of the seven amino acid substitutions of xyn-CDBFV xylanase (E24V, S128F, and G214V) raised the degree of hydrophobicity of the enzyme, whereas two others (V218A and I97S) lowered it. It appears that the degree of hydrophobicity at some specific sites may determine the degree of alkalophilicity of the xylanase. The N164D substitution increases the negative charge of the xylanase, which may also help increase alkalophilicity. Russell and Fersht (1987) described how an increase in the negative charge on

the surface of the model enzyme subtilisin raised the pK_a values of acidic groups and how, with the increased pK_a of the catalytic acidic amino acid, the enzyme can shift its pH optimum to alkaline values. Judging from amino acid sequence alignment from other xylanases (Gruber et al. 1998), the two acidic amino acids E117 and E210 (Fig. 1) are the catalytic sites of xyn-CD/WT. As well, the K168N substitution can raise the negative charge indirectly by removing the positively charged lysine, thus acting cooperatively to raise the alkalophilicity of the xylanase together with other substitutions. Thus, the observations made in the present study are consistent with the ideas presented by Russell and Fersht (1987).

Temperature optima of xyn-CD/WT and xyn-CDBFV

Thermophilicity and thermotolerance of xyn-CDBFV were greatly enhanced as compared with its parent enzyme xyn-CD/WT. The composite mutant enzyme exhibited maximal activity at 62°C compared with the temperature optimum of 47°C determined for the wild-type xyn-CD/WT (Fig. 4). At 60°C, the specific activity of xyn-CDBFV (13 166 μmol glucose·min^{-1}·(mg protein)$^{-1}$) was 130% of its activity at 47°C (Table 1; Fig. 4), and at 77°C, xyn-CDBFV still maintained approximately 10% of its maximal activity. By comparison, the maximal activity of xyn-CD/WT recorded at 47°C (4343 μmol glucose·min^{-1}·(mg protein)$^{-1}$) was reduced to 2% at only 57°C.

Assays for thermostability revealed that maximal activities of both xylanases were exhibited after a 10-min incubation at 60°C (Fig. 5). The activity of xyn-CD/WT was reduced to 50% of its maximum after only 60 min at 60°C, whereas xyn-CDBFV retained 50% of its activity for 78 min. Thus, xyn-CDBFV is slightly more thermostable at 60°C (at pH 6.0) than is its wild-type parent, xyn-CD/WT. At 55°C, both xyn-CDBFV and xyn-CD/WT exhibited over 90% of their maximal activities, even after 120 min of incubation (data not shown).

Resistance of xyn-CDBFV to alkaline conditions

The specific activity of xyn-CDBFV increased to 150% of its 0-h activity (67 μmol glucose·min^{-1}·(mg protein)$^{-1}$) during the initial 2 h of incubation at pH 10.0 (Fig. 6), and this enhanced activity was maintained for at least 24 h. Thus, the alkaline stability of xyn-CDBFV at pH 10.0 was clearly demonstrated. Comparisons with xyn-CD/WT could not be made, however, because the latter enzyme exhibited no xylanolytic activity at that pH (data not shown). Consequently, specific contributions of the seven mutations to the property of alkaline stability could not be elucidated.

In conclusion, the techniques used in this study (error-prone PCR to generate mutants, agar overlay to screen for desired activities, DNA sequencing to identify sites and mutations, and site-directed mutagenesis to create composite mutants) were extremely effective for modifying a thermophilic (half time at 60°C was 60 min) xylanase from *N. patriciarum* for use in paper pulp bio-bleaching. The final composite mutant (xyn-CDBFV) maintained a similar proportion of activity as compared with the parent enzyme at pH 4.0–6.5, but the proportion of activity it retained at pH 7.5–9.0 was much higher than that of the wild type. The enzyme was stable at pH 10.0 for at least 24 h, and its temperature optimum at pH 6.0 was shifted from 47°C (wild type) to 62°C. The specific activity of the composite enzyme was about 2.3 times that of the wild type at pH 6.0. Whereas xyn-CD/WT was nonfunctional at pH 9.0, xyn-CDBFV exhibited specific activity of over 2500 μmol glucose·min^{-1}·(mg protein)$^{-1}$. At 60°C, composite xyn-CDBFV maintained its activity longer than did xyn-CD/WT. These properties render xyn-CDBFV greatly improved over xyn-CD/WT for use in bio-bleaching, where temperatures of 50 or 65°C and alkalinity of pH 7.0 or 9.0 would be encountered (Shoham et al. 1992; Bim and Franco 2000; Georis et al. 2000).

The success of this study demonstrates that a high pH agar overlay is a powerful screening technique that may be applied to other xylanases and other enzymes. Modifications of naturally occurring thermophilic enzymes could enhance their usefulness in a range of industrial applications not limited to the paper pulp industry. Some of the present observations suggest that the degree of hydrophobicity at some specific sites of the xylanases may determine their alkalophilicity. Further research is necessary to better define the relationships between these characteristics, in this and other xylanases or other enzymes.

Acknowledgements

The authors thank Dr. G. G. Chang of the Department of Biochemistry, National Defense Medical Center, Republic of China for critical review of the manuscript. This study was partially funded by the Taiwan Sugar Corporation (Grant NSC89-TSC-7-001-003) and by the Council of Agriculture Republic of China (Grant 89-2101-24-01-03-05).

References

Arase, A., Yomo, T., Urabe, I., Hata, Y., Katsube, Y., and Okada, H. 1993. Stabilization of xylanase by random mutagenesis. FEBS Lett. **316**: 123–127.

Bajpai, P. 1999. Application of enzymes in the pulp and paper industry. Biotechnol. Prog. **15**: 147–157.

Biely, P. 1985. Microbial xylanolytic systems. Trends Biotechnol. **3**: 286–290.

Bim, M.A., and Franco, T.T. 2000. Extraction in aqueous two-phase systems of alkaline xylanase produced by *Bacillus pumilus* and its application in kraft pulp bleaching. J. Chromatogr. B, **743**: 349–356.

Buchert, J., Tenkanen, M., Kantelinen, A., and Viikari, L. 1994. Application of xylanases in the pulp and paper industry. Bioresourc. Technol. **50**: 65–72.

Cha, J., and Batt, C.A. 1998. Lowering the pH optimum of D-xylose isomerase: the effect of mutations of the negatively charged residues. Mol. Cells, **8**: 374–382.

Georis, J., Giannotta, F., Buyl, E.D., Granier, B., and Frere, J.M. 2000. Purification and properties of three endo-β-1,4-xylanases produced by *Streptomyces* sp. strain S38 which differ in their ability to enhance the bleaching of kraft pulps. Enzyme Microb. Technol. **26**: 178–186.

Giver, L., Gershenson, A., Freskgard, P.O., and Arnold, F.A. 1998. Directed evolution of a thermostable esterase. Proc. Natl. Acad. Sci. U.S.A. **95**: 12 809 – 12 813.

Gruber, K., Klintschar, G., Hayn, M., Schlacher, A., Steiner, W., and Kratky, C. 1998. Thermophilic xylanase from *Thermomyces*

lanuginosus: high-resolution X-ray structure and modeling studies. Biochemistry, **37**: 13 475 – 13 485.

Harris, G.W., Pickersgill, R.W., Connerton, I., Debeire, P., Touzel, J.P., Breton, C., and Perez, S. 1997. Structural basis of the properties of an industrially relevant thermophilic xylanase. Proteins, **29**: 77–86.

Kulkarni, N., Shendye, A., and Rao, M. 1999. Molecular and biotechnological aspects of xylanases. FEMS Microbiol. Rev. **23**: 411–456.

Liu, J.H., Selinger, B.L., Tsai, C.-F., and Cheng, K.-J. 1999. Characterization of a *Neocallimastix patriciarum* xylanase gene and its product. Can. J. Microbiol. **45**: 970–974.

Maat, J., Roza, M., Verbakel, J., Stam, H., Santos da Silva, M.J., Bosse, M., Egmond, M.R., Hagemans, M.L.D., v. Gorcom, R.F.M., Hessing, J.G.M., v.d. Hondel, C.A.M.J.J., and v. Rotterdam, C. 1992. Xylanases and their application in bakery. *In* Xylans and xylanases. *Edited by* J. Visser, M.A. Kusters van Someren, G. Beldman, and A.G.J. Voragen. Elsevier Science Publishers B.V., Oxford, U.K. pp. 349–360.

Miller, J.H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Nishiya, Y., Harada, N., Teshima, S.I., Yamashita, M., Fuji, I., Hirayama, N., and Murookam Y. 1997. Improvement of thermostability of *Streptomyces* cholesterol oxidase by random mutagenesis and a structural interpretation. Protein Eng. **10**: 231–235.

Russell, A., and Fersht, A.R. 1987. Rational modification of enzyme catalysis by engineering surface charge. Nature (London), **328**: 496–500.

Shoham, Y., Schwartz, Z., Khasin, A., Gat, O., Zosim, Z., and Rosenberg, E. 1992. Delignification of wood pulp by a thermostable xylanase from *Bacillus stearothermophilus* strain T-6. Biodegradation, **3**: 207–218.

Simpson, H.D., Haufler, U.R., and Daniel, R.M. 1991. An extremely thermostable xylanase from the thermophilic eubacterium *Thermotoga*. Biochem. J. **277**: 413–417.

Song, J.K., and Rhee, J.S. 2000. Simultaneous enhancement of thermostability and catalytic activity of phospholipase A1 by evolutionary molecular engineering. Appl. Environ. Microbiol. **66**: 890–894.

Tamblyn Lee, J.M., Hu, Y., Zhu, H., Cheng, K.-J., Krell, P.J., and Forsberg, C.W. 1993. Cloning of a xylanase gene from the ruminal fungus *Neocallimastix patriciarum* 27 and its expression in *Escherichia coli*. Can. J. Microbiol. **39**: 134–139.

Teather, R.M., and Wood, P.J. 1982. Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl. Environ. Microbiol. **43**: 777–780.

Uhlig, H. 1998. Industrial enzymes and their applications. John Wiley & Sons Inc., New York. p. 394. (Translated from German and updated by E.M. Linsmaier-Bednar.)

Viikari, L., Kantelinen, A., Sundquist, J., and Linko, M. 1994. Xylanases in bleaching: from an idea to the industry. FEMS Microbiol. Rev. **13**: 335–350.

Zhao, H., and Arnold, F.H. 1999. Directed evolution converts subtilisin E into a functional equivalent of thermitase. Protein Eng. **12**: 47–53.

The EMBO Journal vol.13 no.11 pp.2493–2501, 1994

Three-dimensional structure of endo-1,4-β-xylanase II from *Trichoderma reesei*: two conformational states in the active site

Anneli Törrönen, Anu Harkki[1] and Juha Rouvinen[2]

Department of Chemistry, University of Joensuu, PO Box 111, FIN-80101 Joensuu, Finland and [1]Cultor Ltd Technology Center, FIN-02460 Kantvik, Finland
[2]Corresponding author

Communicated by J.Janin

The three-dimensional structure of endo-1,4-β-xylanase II (XYNII) from *Trichoderma reesei* has been determined by X-ray diffraction techniques and refined to a conventional *R*-factor of 18.3% at 1.8 Å resolution. The 190 amino acid length protein was found to exist as a single domain where the main chain folds to form two mostly antiparallel β-sheets, which are packed against each other in parallel. The β-sheet structure is twisted, forming a large cleft on one side of the molecule. The structure of XYNII resembles that of *Bacillus* 1,3–1,4-β-glucanase. The cleft is an obvious suggestion for an active site, which has putative binding sites for at least four xylose residues. The catalytic residues are apparently the two glutamic acid residues (Glu86 and Glu177) in the middle of the cleft. One structure was determined at pH 5.0, corresponding to the pH optimum of XYNII. The second structure was determined at pH 6.5, where enzyme activity is reduced considerably. A clear structural change was observed, especially in the position of the side chain of Glu177. The observed conformational change is probably important for the mechanism of catalysis in XYNII.

Key words: conformational change/crystal structure/*Trichoderma*/xylanase

Introduction

Xylan is the most abundant hemicellulose in plants, accounting for as much as 35% of the dry weight of higher plants. Xylans are heteropolysaccharides, the backbone of which is composed of ~150–200 β-1,4-linked xylopyranose units. The degree of substitution of xylan varies, depending on its botanical origin. The substituents, like arabinose, 4-*O*-Me-D-glucuronic acid and acetic acid, are bound covalently to the xylopyranose units. The action of several hemicellulolytic enzymes is needed to hydrolyze xylan completely. During the last few years, the enzymatic hydrolysis of xylan has raised significant interest because of its applications, for example, in biobleaching and the food and feed industry (Biely, 1985; Poutanen *et al.*, 1987; Viikari *et al.*, 1990; Wong and Saddler, 1992; Gilbert and Hazlewood, 1993).

The enzymes acting on the backbone of xylans are exo-1,4-β-xylosidases (EC 3.2.1.37) and endo-1,4-β-xylanases (EC 3.2.1.8). Endoxylanases are capable of hydrolyzing the internal 1,4-β-bonds of the xylan backbone and thereby produce several xylo-oligomers of varying length. About 20 amino acid sequences of different low-molecular-weight xylanases of bacterial and fungal origin have been published. The comparison of available β-glycanase sequences by hydrophobic cluster analysis (HCA) has revealed eight families (A–H) (Gilkes *et al.*, 1991; Henrissat, 1991; Törrönen *et al.*, 1993a). So far, X-ray structures of five glycosidic enzymes have been reported: the catalytic domain of cellobiohydrolase II from *Trichoderma reesei* (family B) (Rouvinen *et al.*, 1990), the catalytic domain of endoglucanase E2 from *Thermomonospora fusca* (family B), endoglucanase CelD from *Clostridium thermocellum* (family E) (Juy *et al.*, 1992), endoglucanase V from *Humicola insolens* (Davies *et al.*, 1993) and *Bacillus* H(A16-M) hybrid 1,3–1,4-glucanase (Keitel *et al.*, 1993).

Low-molecular-weight xylanases belong to the β-glycanase family G, and their amino acid sequences are highly conserved (Törrönen *et al.*, 1993a). The crystallization of seven different low-molecular-weight xylanases has been reported (Moriyama *et al.*, 1987; Rose *et al.*, 1987; Golubev *et al.*, 1993; Pickersgill *et al.*, 1993; Törrönen *et al.*, 1993b; Viswamitra *et al.*, 1993). However, the only structural information published so far concerns the folding pattern of IPO endo-1,4-β-xylanase from *Bacillus pumilus* (Arase *et al.*, 1993).

The filamentous fungus *T.reesei* is known to be one of the most effective producers of polysaccharide hydrolyzing enzymes (Kubicek *et al.*, 1993). The two major inducible homologous endo-β-1,4-xylanases (EC 3.2.1.8) secreted by *T.reesei* are XYNI and XYNII (Törrönen *et al.*, 1992). They are both small protein molecules with molecular masses of 19 and 21 kDa, and consisting of 178 and 190 amino acids, respectively. The isoelectric point is 5.2 for XYNI and 9.0 for XYNII. The pH optimum and product distribution for XYNI and XYNII are different. XYNII alone represents >50% of the xylanolytic activity of the culture filtrate of *T.reesei* grown on xylan. We have recently crystallized both enzymes (Törrönen *et al.*, 1993b) and now report the three-dimensional structure of XYNII.

Results

Crystallography

The three-dimensional structure of XYNII has been solved using the multiple isomorphous replacement method augmented with anomalous scattering information. The final model contains two xylanase molecules (called A and B) in an asymmetric unit. All the 190 amino acid residues and 1480 atoms for both enzyme molecules are included in the model. The first residue in both molecules is pyrrolidone carboxylic acid. The number of water molecules in the final model is 344 (0.9/residue). The final model contains a total of 3304 atoms. The r.m.s. fit between the two molecules,

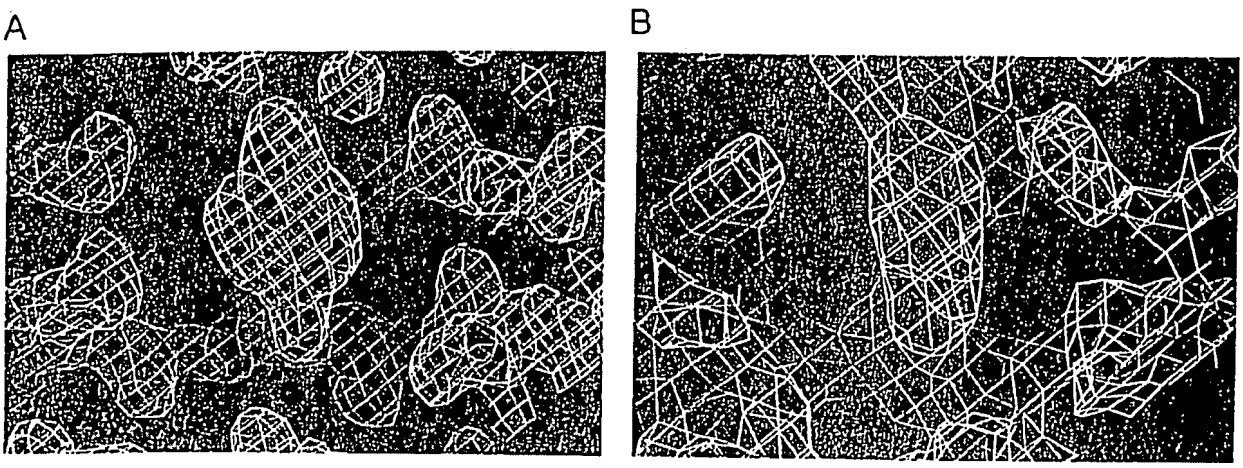


Fig. 1. (A) A portion of the final model with electron density (coefficients $2F_{obs}-F_{calc}$, α_{calc}, contoured at 1σ level) around Tyr87. (B) The same region of the original (unaveraged) 2.8 Å MIR map with the refined model.

A and B, is 0.231 Å (calculated from the positions of Cα atoms). The *R*-factor is 18.3% for 24 911 reflections with $F \geq 1\sigma(F)$ between 8.0 and 1.8 Å (75% complete). The overall geometry is very good (0.008 Å r.m.s. distances, 2.5° r.m.s. angles and 0.9° r.m.s. for fixed dihedrals). The real-space correlation coefficient (Jones *et al.*, 1991), calculated for the final model using all atoms as a function of residue, showed good values. The coefficient was smallest for residues 164–165 in molecule A (0.64). In molecule B, these residues had a much stronger electron density. Asp140 was the only residue which fell outside the allowed regions on the Ramachandran map. A portion of the final electron density map together with the corresponding MIR map is shown in Figure 1.

The second model for XYNII was calculated using the data from crystal soaked in a solution containing xylose at pH 6.5. The final model contains 3304 atoms. The *R*-factor is 18.4% for 17 684 reflections with $|F| \geq 1\sigma(F)$ between 8.0 and 1.8 Å (53.7% complete, 84.1% complete at 2.5 Å resolution). The average deviation from the ideal in bond lengths is 0.009 Å, in bond angles 2.7° and in fixed dihedral angles 0.9°. The r.m.s. difference between the two XYNII models is 0.148 Å and was calculated from the positions of all Cα atoms in the molecule A.

Molecular structure

The schematic drawing and naming of the secondary structure elements are presented in Figure 2A. The overall shape of this globular protein is ellipsoidal, having approximated dimensions of 32 × 34 × 42 Å. XYNII has 190 amino acids; it is a single-domain polypeptide containing two β-sheets (A and B) and one α-helix (Figure 3). The structure contains a total of 15 β-strands. All β-strands have antiparallel hydrogen bonding, except B6b and B7, which have mixed bonding. Both β-sheets A and B twist and form a cleft on one side of the protein. The hydrophobic faces of these β-sheets are packed against each other in parallel in a sandwich-like manner, forming the hydrophobic core of the protein. The hydrophilic face of β-sheet A forms a flat surface with many serine and threonine residues, and is accessible to solvent. The hydrophilic face of β-sheet B makes a surface for the cleft.

β-Sheet A is composed of five antiparallel β-strands. The

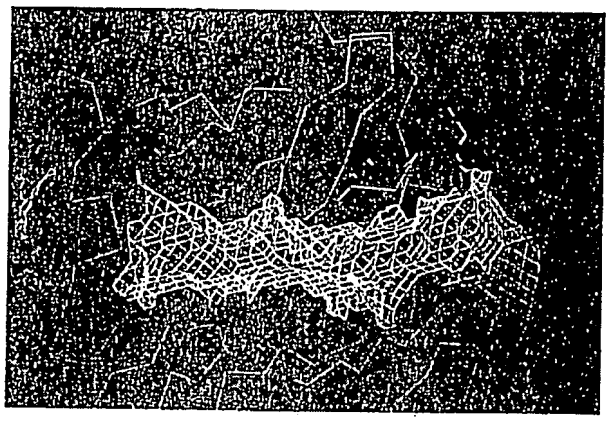


Fig. 4. Solvent-accessible surface of the cleft in XYNII structure. Two catalytic residues Glu86 and Glu177 are shown.

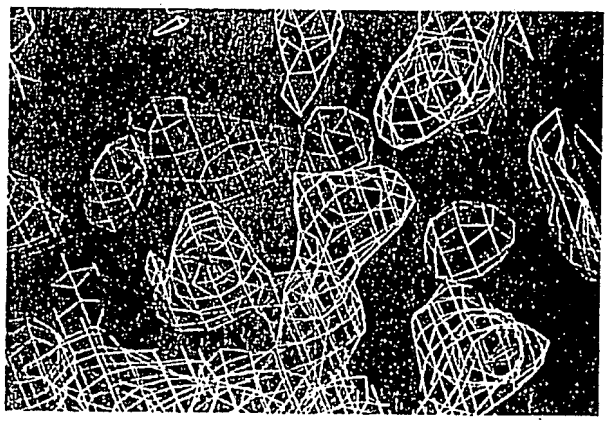


Fig. 5. The observed [illegible] in Glu177 position. The electron density coefficients $2F_{obs}-F_{calc}$, α_{calc}, contoured at 1σ levels around Glu177 for a pH 5.0 [illegible] pH 6.5 model (right).

putative β strand A1 is missing in XYNII, but may exist in some other homologous xylanases. Strands A2, A3 and A6 are shorter (four or five residues) than the middle strands A4 and A5 (nine and 11 residues). The short strands are quite planar, whereas the long strands are twisted. Four

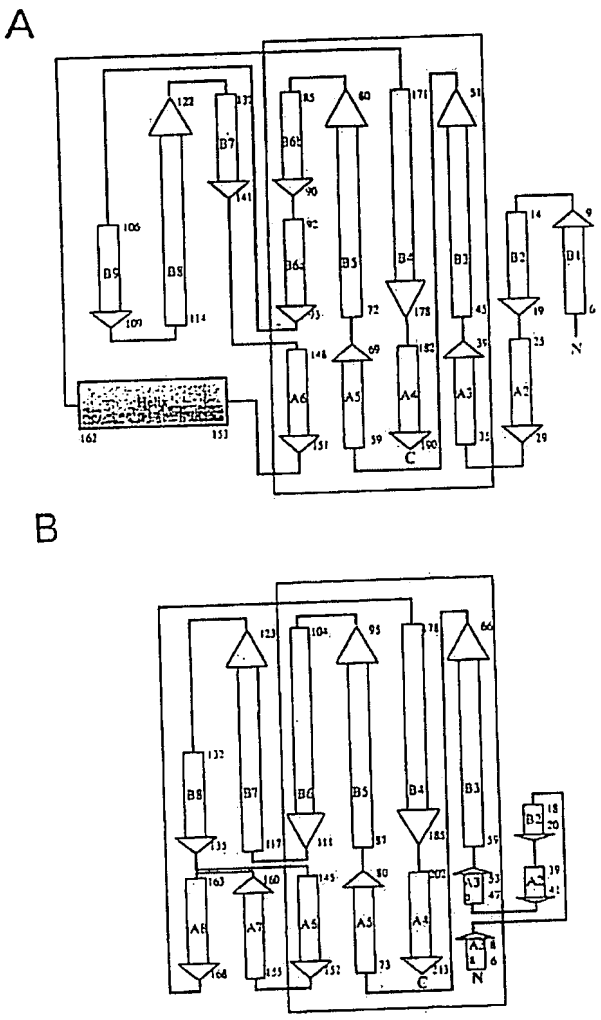


Fig. 2. The schematic drawing of the secondary structure elements (A) XYNII from *T.reesei* and (B) *Bacillus* 1,3−1,4-β-glucanase. The arrows mark β-strands and cylinder α-helix. The structure elements of 1,3−1,4-β-glucanase are adopted from the work of Keitel *et al.* (1993), from which the small helix between strands B4 and A4 is omitted for clarity. The corresponding elements between two proteins are numbered in a similar way. The common central folding pattern in both proteins is boxed.

strands are connected at the same end of β-sheet A to β-sheet B. These loops have different lengths (2−5 residues). The connection between A6 and B6 is interrupted by a long insertion.

β-Sheet B contains β-strands B1−B9. Strands B1, B2 and B9 are short (four, six and four residues), and B3, B4, B5 and B8 are longer (seven, eight, nine and nine residues). One residue divides unit B6 into two shorter strands, B6a and B6b, which are twisted −90°. There is also a twist of −90° in the middle of the long β-strands B3, B4 and B5. The twisted part of β-sheet B has been described as a separate sheet in xylanase from *B.pumilus* (Arase *et al.*, 1993). The insertion between strands B6 and A6 contains three β-strands: B7, B8 and B9. The only α-helix is located between β-strands A6 and B4, and is packed against the hydrophobic face of β-sheet B.

The overall structure has the shape of a 'right hand' (Figure 3A and B), where the two β-sheets (A and B) form 'fingers' and the twisted part of β-sheet B and the α-helix form a 'palm' (Figure 3A). The long loop of the nine residues between β-strands B7 and B8 makes a 'thumb', which partly closes the cleft. A striking feature is a 12 residue long irregular loop between β-strands B6a and B9. A part of this, residues 96−102 (Tyr−Asn−Pro−Ser−Thr−Gly−Ala), form a 'cord', which partly closes the cleft on one side. The residues in the cord have a clear electron density, indicating that this unit has a well-defined structure. No hydrogen bonds between the cord and the rest of the molecule were observed, but some hydrophobic contacts were obvious.

XYNII contains no cysteine residues, so there are no disulfide bridges. However, there are eight negatively charged (aspartate and glutamate) residues and 13 positively charged (lysine, arginine and histidine) residues in the protein molecule. These form ionic bridges: between Asp170 and Arg81, Glu91 and Arg141/Arg145, Glu107 and Lys104/Arg142, and Asp116 and Lys104/Arg142. None of these residues is completely conserved. Lys49 has an important structural role. It places a positively charged nitrogen in a position inside the protein and makes hydrogen bonds with the main chain carbonyl oxygens of Asn10, Gly30, Gly32 and Gln34.

The active site cleft and conformational change

The shape of the cleft was estimated by using the solvent-accessible surface (Voointholt *et al.*, 1989) (Figure 4). The active site is likely to be located in this cleft, which has a length and depth of ~25 and 9 Å, respectively. The average width of the cleft is ~4 Å, but there are two regions where the width is smaller: one between Trp18 and Pro126 in the middle of the cleft, and the other between Tyr96 and Tyr179 at the end of the cleft.

XYNII was crystallized at pH 5.0, which also corresponds to the pH optimum of the enzyme (Tenkanen *et al.*, 1992; Törrönen *et al.*, 1992). In order to get a complex structure, crystals were soaked in a solution containing xylose at pH 6.5. At this pH, the activity of XYNII is considerably reduced and xylose binding was expected to be better. Clear structural changes for some residues in the cleft were observed in both molecules A and B (Figure 5). Unfortunately, no clear electron density was found for the ligand, probably due to the low binding affinity of xylose. Later, the effect of pH was studied by collecting a new data set at 2.0 Å resolution at pH 6.5, but now without xylose (data not shown). Again, similar conformational change was observed.

The largest change was observed for Glu177, where the torsion angle between Cβ and Cγ atoms was changed ~100° (from −89° to 171°). This places the carboxyl part of the side chain in a different position, the displacement being 2.8 Å calculated from the positions of the Cδ atom. In addition, a change in the conformation of the side chain of Tyr73 and Tyr88 can be detected. The bond between the atoms Cα and Cβ of Tyr73 rotates ~78° (from 55° to −23°), and the bond between the atoms Cβ and Cγ of Tyr73 −47° (from 82° to 129°). Both rotations cause a change of ~4.0 Å in the position of the Oη atom of Tyr73. The position of the side chain of Tyr88 is less affected. The bond between the atoms Cα and Cβ of Tyr88 rotates ~9° (from −58° to −67°) and the bond between the atoms Cβ and Cγ of Tyr88 −12° (from 83° to 71°). These rotations lead to an ~0.9

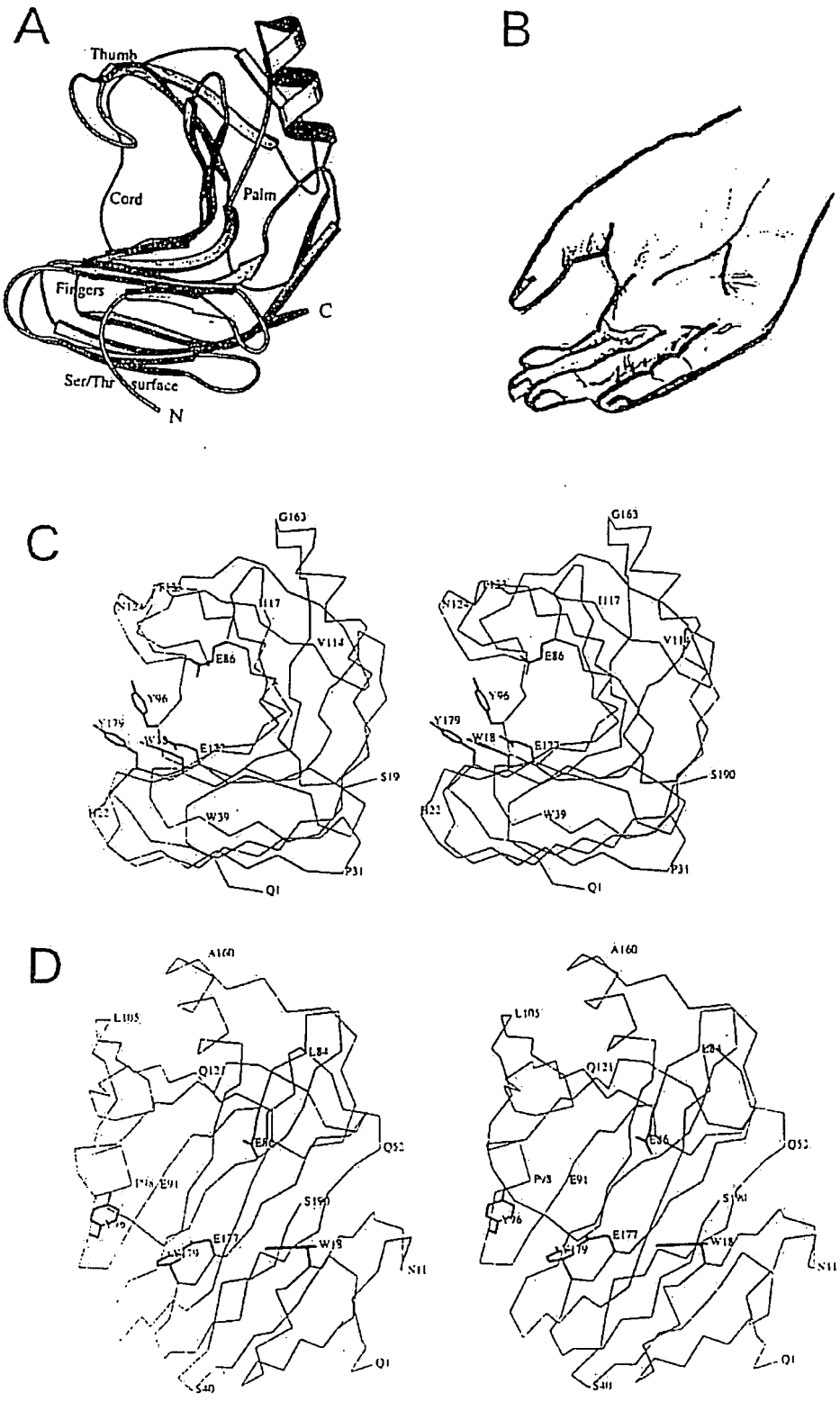


Fig. 3. (A) Ribbon representation of the XYNII molecule showing an α-helix and β-strands. The structure is reminiscent of the shape of a 'right hand' (B) and this analogy is shown. The picture was generated with the MolScript program (Kraulis, 1991). (C) The Cα skeleton of XYNII in stereo. The putative catalytic residues (Glu86 and Glu177) and aromatic binding residues (Tyr96, Tyr179 and Trp18) are shown. (D) As in (C), but rotated along the *y*-axis −90°.

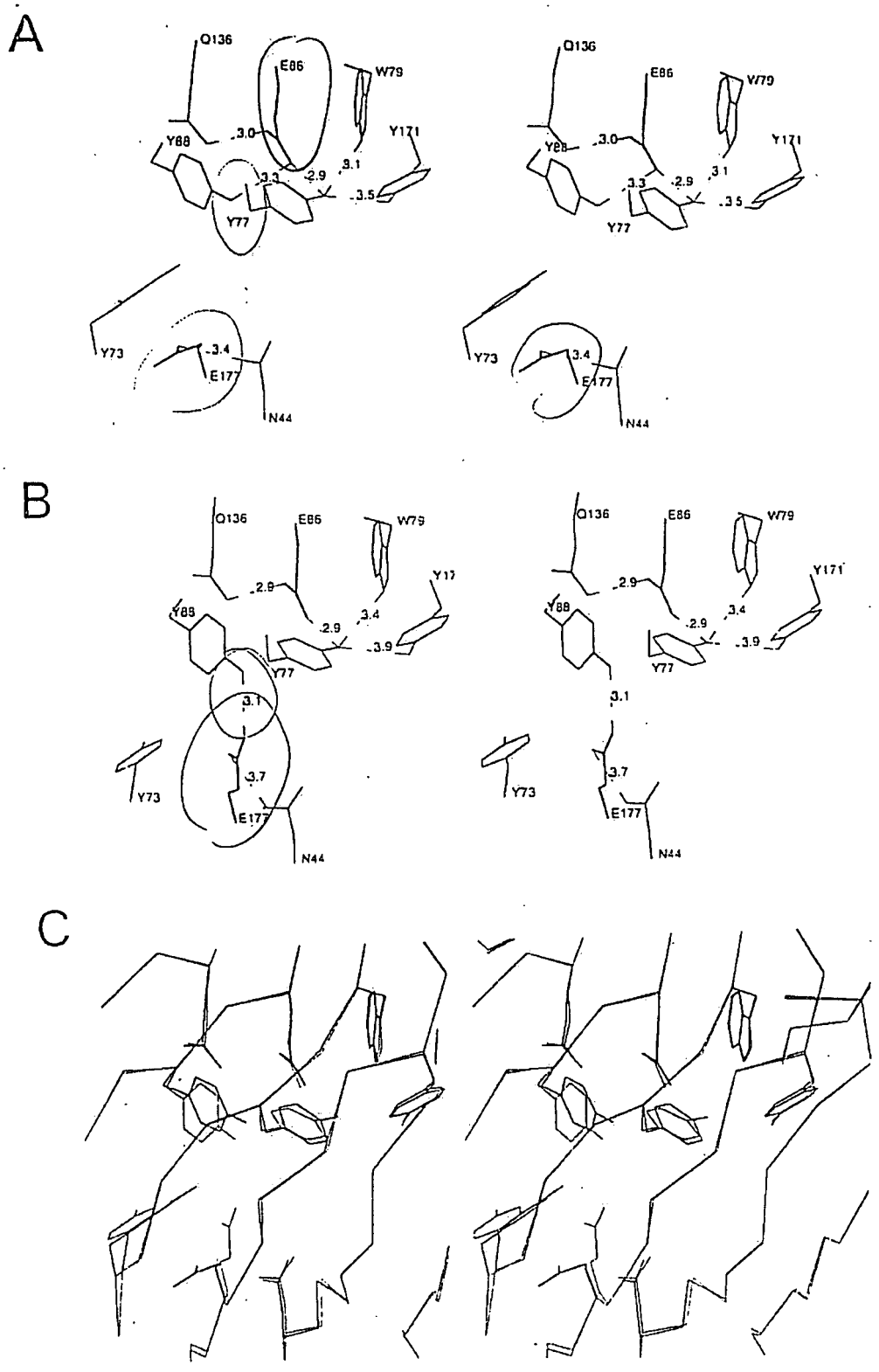


Fig. 6. Hydrogen bond network around Glu86 and Glu177 at (A) pH 5.0, (B) pH 6.5 and (C) both together with the main Cα chain.

Å change in the position of Oη of Tyr88. In spite of the ~0.6 Å change in the position of the 'cord', the differences between the two structures elsewhere in the protein molecule were small, close to the experimental errors.

The hydrogen bonds of Glu86 and Glu177

In previous structural studies it has been suggested that Glu93 and Glu182 are the catalytic residues in *B.pumilus* IPO xylanase (corresponding to Glu86 and Glu177 in XYNII)

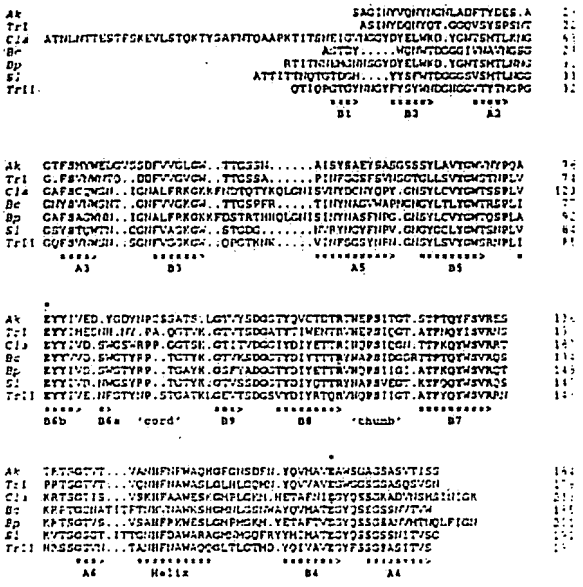


Fig. 7. Multiple sequence alignment among endo-1,4-β-xylanases (family G). The totally conserved amino acids are in bold-face type and the stars indicate the putative catalytic residues. The secondary structure assignment of XYNII (*TrII*) is indicated by arrows (see Figure 2). The positions of the 'thumb' and 'cord' are marked. *Ak*, *Aspergillus kawachii* (Ito *et al.*, 1992); *Tr1*, *Trichoderma reesei* XYNI (Törrönen *et al.*, 1992); *Cla*, *Clostridium acetobutylicum* (Zappe *et al.*, 1990); *Bc*, *Bacillus circulans* (Yang *et al.*, 1988); *Bp*, *Bacillus pumilus* (Fukusaki *et al.*, 1984); *Sl*, *Streptomyces lividans* (Shareck *et al.*, 1991); *TrII*, *Trichoderma reesei* XYNII (Törrönen *et al.*, 1992).

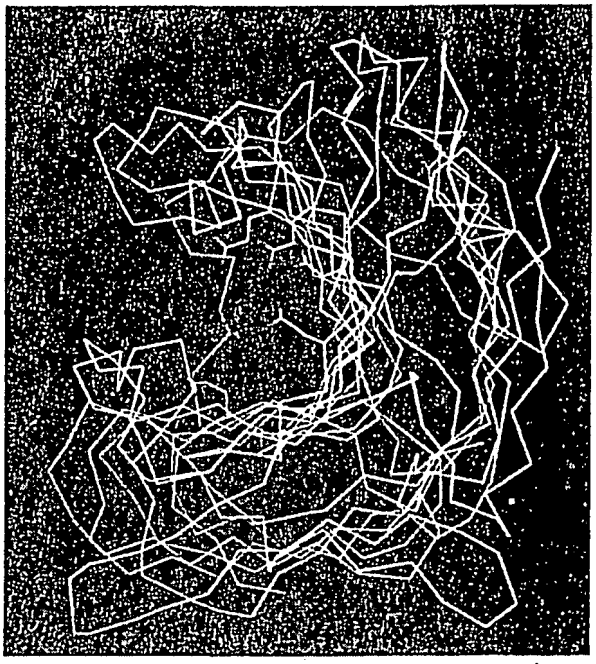


Fig. 8. The superimposition of XYNII from *T.reesei* (green) and 1,3–1,4-β-glucanase from *Bacillus* (yellow).

(Ko *et al.*, 1992). The conformational change in XYNII observed in this study alters the hydrogen bonding pattern of both residues (Figure 6). In the pH 5.0 structure, the Oε1 of Glu177 forms a hydrogen bond to the Nδ2 of Asn44, but the Oε2 does not form any hydrogen bonds. Thus, it is possible that Oε2 has a proton and Glu177 is neutral in this conformation. In the other (pH 6.5) structure, the position of the carboxyl group shifted. There Oε1 still forms a weak hydrogen bond to Nδ2 of Asn44, whereas strong hydrogen bondings to Oη of Tyr88 and to one water molecule were observed. Glu177 is likely to be in a charged form in this second observed conformation (pH 6.5), because both its Oε1 and Oε2 accept hydrogen bonds.

No clear conformational differences were detected in the position of Glu86 in these two structures. However, some changes were observed in the hydrogen bonding of Glu86. In the native structure, Oε1 of Glu86 makes a strong hydrogen bond to Nε2 of Gln136. Oε2 forms one hydrogen bond to Oη of Tyr 77 and one to Oη of Tyr88. The hydrogen bond network continues from Oη of Tyr77 to Nε1 of Trp79 and to Oη of Tyr171. In the other (pH 6.5) structure, the distance between Oε2 of Glu86 and Oη of Tyr88 increases from 3.3 to 4.1 Å, because of the position shift of 0.9 Å of Oη of Tyr88, and the hydrogen bonding disappears. To compensate for this loss, the hydrogen bond between Oε2 of Glu86 and Nε2 of Gln136 shortens from 3.0 to 2.9 Å. The link between the putative catalytic residues Glu86 and Glu177 is a hydroxyl group of Tyr88. The proton attached to Oη of Tyr88 points towards Glu86 in 'pH 5.0' conformation and towards Glu177 in 'pH 6.5' conformation.

Discussion

Family G xylanases

Glycanase family G was first described by Henrissat (1991). Recently, hydrophobic cluster analysis was used to analyze the amino acid sequences of 17 different low-molecular-weight endo-β-1,4-xylanases (Törrönen *et al.*, 1993a). The amino acid homology in this family is high, indicating a similar fold. The alignment of seven amino acid sequences in this family is shown in Figure 7, together with the secondary structure assignment of XYNII. XYNII has a total of 32 residues (17%), which are common in all these proteins. Both the putative catalytic residues Glu86 and Glu177 of XYNII are conserved. There is a clear cluster of conserved residues around Glu86. The direct hydrogen bonds to Glu86 (from Gln136, Tyr77 and Tyr88) are conserved, as is a second 'layer' hydrogen bond from Trp79 to Tyr77. The hydrogen bond from Tyr171 to Tyr77 also exists in other xylanases, although the tyrosine residue is replaced by histidine in some members of the family. This cluster of five residues around Glu86 forms a hydrogen bond network, which is capable of 'fine tuning' the properties of Glu86.

The situation is different around Glu177, where the surrounding is much less conserved. If Glu177 initiates the reaction by donating a hydrogen, the pK_a value of its ionizable side chain directly affects the catalytic properties of the whole enzyme. The pK_a value for 'free' glutamic acid is 4.6–5.0. XYNII has been reported to be most active in a higher pH range: 5.0–5.5 (Tenkanen *et al.*, 1992; Törrönen *et al.*, 1992). In other xylanases, the pH optimum has been estimated to vary between 3.5 and 7.0 (Wong *et al.*, 1992). The variations are probably due to the different amino acid residues in the neighborhood of Glu177.

Many of the conserved residues of endo-1,4-β-xylanases are believed to be structurally important in order to confirm

Table I. Data collection and MIR statistics

Crystal and soaking data

Data set	Soaked compound	pH	Soaking conc. (mM)	Soaking time (h)	Unit cell dimensions			
					a (Å)	*b* (Å)	*c* (Å)	β (°)
Native		5.0			81.57	60.63	38.25	94.4
PT1	$K_2Pt(CN)_4$	5.0	50	17	81.69	60.82	38.10	94.5
PT2	$K_2Pt(CN)_4$	5.0	50	17	81.51	60.72	38.05	94.4
HG1	$Hg(OAc)_2$ + thioxylose	5.0	20	20	81.71	60.96	38.40	94.4
HG2	$Hg(OAc)_2$ + thioxylose	8.0	20	10	81.35	60.67	38.23	94.4
HG3	MeHgOAc + thioxylose	5.0	20	18	81.47	60.72	38.16	94.4
Native2	xylose	6.5	10	22	81.70	60.96	38.05	94.3

Diffraction data

Data set	dmin (Å)	No. of measurements	No. of unique reflections	Completeness of data (%)	R_{merge} (%)
Native	1.8	84 025	25 322	75	4.95
PT1	2.0	77 296	21 886	86	7.31
PT2	2.5	38 961	12 114	93	7.63
HG1	2.5	43 235	12 336	94	4.91
HG2	2.5	29 763	11 047	85	10.74
HG3	2.0	45 605	19 484	77	8.42
Native2	1.8	44 107	18 071	54	8.77

MIR analysis

	Occupancy	X	Y	Z	R.m.s.*Fh*/residual	R_{Cullis}
PT1 + PT2	31.01	0.620	0.195	0.102	1.15	0.858
	23.64	0.144	0.757	0.421		
	10.51	0.592	0.185	0.413		
	9.07	0.089	0.751	0.109		
HG1	18.23	0.197	0.000	0.111	1.53	0.723
	20.93	0.694	0.942	0.454		
	8.97	0.588	0.536	0.272		
	8.05	0.080	0.396	0.262		
HG2	27.59	0.198	0.009	0.113	1.77	0.608
	30.64	0.694	0.948	0.545		
	18.05	0.589	0.541	0.272		
	17.55	0.078	0.404	0.264		
HG3	27.11	0.269	0.509	0.442	0.99	0.808
	23.72	0.237	0.945	0.022		

Mean figure of merit = 0.50

where $R_{merge} = 100 \times \Sigma\Sigma |F^2(i) - \langle F^2(i) \rangle| / \Sigma\Sigma F^2(i)$, R.m.s.*Fh*/residual = $[(\Sigma F_H^2)/(\Sigma F_{DER} - F_{PH})^2]^{1/2}$, where F_H is the heavy atom form factor, F_{DER} and F_{PH} are the derivative structure factor and calculated structure factor. $R_{Cullis} = \Sigma(|F_{DER} - F_{PH}|)/\Sigma(|F_{DER} - F_{NAT}|)$, where F_{NAT} is the native structure factor and the summation is taken over the centric reflections.

its position. This feature could be useful in catalysis, where the glutamic acid residue has a dual role. First, it will initiate the reaction by donating a proton to the substrate; second, it may assist (in a charged form) a water molecule to attack the substrate. XYNII has been reported also to have transxylosidase activity (Tenkanen *et al.*, 1992). The second (pH 6.5) conformation is probably involved in that type of catalysis.

Materials and methods

Crystallization and data collection

Purification and crystallization of native endo-1,4-β-xylanase II have been described by Törrönen *et al.* (1992, 1993b). The crystals are monoclinic, belonging to the space group $P2_1$. Intensity data were collected from XYNII crystals on a Rigaku R-AXIS IIC area detector using CuKα radiation produced by a Rigaku RU200HB rotating anode (50 kV, 180 mA). The area detector data were processed using R-AXIS IIC software. Data collection statistics are given in Table I. We deduced from the unit cell volume that there are two xylanase molecules in the asymmetric unit. The self-rotation function showed a sharp peak in the position of Ψ = 90°, ρ = 86° and κ = 180°, indicating the presence of a non-crystallographic two-fold element.

Structure determination

In total, 52 different heavy-atom experiments were performed in order to find suitable derivatives for MIR phasing. The difference Patterson maps were calculated with the XtalView program. Only $K_2Pt(CN)_4$ was found to produce a clear substitution for heavy atoms. Later, thioxylose was synthesized by replacing a hydroxyl group of C1 carbon with a thiol group. The XYNII crystals were first soaked in thioxylose solution and then transferred to a mercury acetate solution. Difference Patterson maps showed clear peaks for mercury in the Harker section. The co-ordinates of heavy atoms were calculated from the Patterson maps and their positions were refined with the Protein program (Steigemann, 1992). The electron density

the correct folding and packing. In addition, there are three residues in which the conservation is interesting. One of them, Pro98 in *trans*-conformation, located in the middle of the 'cord', seems to be structurally important. It also addresses the question of whether or not its conformation is capable of changing from *trans* to *cis*. If this is possible, it could lead to a large structural change in the stucture. Naturally, this hypothesis should be demonstrated experimentally. Two other interesting conserved residues are the Asn124 and Thr133 above the 'thumb' (Figure 3C). The side chains of these two residues have no effect on packing. Both of them point towards solvent without making any important hydrogen bonds to other residues of the protein molecule. The role of this conservation is unclear, although we can speculate that they take part in substrate binding.

The flat Ser/Thr face of β-sheet A is also quite conserved in family G. The function of this face is difficult to understand, but it may take part in the penetration of the enzyme into small pores, which exist in wood components. Its functional role may, to some extent, be similar to separate cellulose-binding domains observed in many cellulases. In addition, it can be noted that the linker region between the cellulase catalytic and substrate-binding domains normally contains numerous serines and threonines, although many of them are believed to be glycosylated (Gilbert and Hazlewood, 1993).

Comparison with Bacillus 1,3–1,4-β-glucanase

The three-dimensional structure of the β-glucan hydrolyzing enzyme, 1,3–1,4-β-glucanase (also called lichenase) from *Bacillus* has recently been reported by Keitel *et al.* (1993). The protein is a hybrid containing residues 1–16 of the mature *Bacillus amyloliquefaciens* enzyme and residues 17–214 from the *Bacillus macerans* enzyme. The general shape of this molecule resembles the XYNII structure, as shown in Figure 8. The r.m.s. fit between the Cα atoms of 83 residues is 2.2 Å, indicating a rather weak structural similarity. The reason for the common general shape, but weak local structural similarity, was found by comparing the topological diagrams of the secondary structure elements (Figure 2). The comparison shows that the central folding pattern (boxed in Figure 2) is similar, indicating the same kind of connection between the central four β-strands. Both structures have an insertion between β-strands B6 and A6. The differences are in the arrangements of the secondary structure elements located outside the central region.

There are three cationic residues, Glu105, Asp107 and Glu109, in the active site cleft of 1,3–1,4-β-glucanase, all having side chains pointing towards the cleft. The study of site-directed mutagenesis has suggested that at least Glu105 is important for catalysis. Interestingly, all three residues are located in β-strand B6, which corresponds to the B6b strand containing Glu86 in the XYNII structure. In comparison, the glucanase structure does not have a corresponding residue for the Glu177 of XYNII.

Implications for catalysis

Without high-resolution complex structures, it is difficult to determine the ligand binding accurately. The soaking experiment with xylose, xylobiose and xylotriose at pH 5.0 and 6.5 did not reveal the binding mode for the ligands. It seems that the binding affinity of xylo-oligomers is too low. For example, Ryazanova *et al.* (1993) have reported that D-xylose binds weakly to different *Aspergillus japonicus* xylanases (K_i = 0.22–1.5 mM).

However, by analyzing the conserved residues in the cleft, some assumptions can be made. In this respect, the most interesting residues are the aromatic residues, which have the hydrophobic face of the side chain on the surface of the cleft. In many carbohydrate-binding proteins, tyrosine, tryptophan or phenylalanine residues participate in the binding of a carbohydrate residue by packing against a flat face of the carbohydrate ring (Vyas, 1991). In the cleft of XYNII, three residues of this kind exist: Trp18, Tyr96 and Tyr179 (Figure 3C and D). Trp18 is located on β-strand B2 at the entrance of the cleft, Tyr179 on β-strand A4 in the middle of the cleft and Tyr96 is located on the 'cord', quite near Tyr179, but on the other side. By using the molecular model of the xylo-oligomer, we can estimate that there is space for at least four xylose units in the cleft (subsites which we describe as A, B, C and D). It is probable that Tyr96 and Tyr179 are part of adjacent subsites (A and B) for xylose residues, whereas Trp18 is part of the fourth subsite (D) further away. Between residues Tyr179 and Trp18 there is space for one xylose residue (subsite C). In addition, we can note that one face of the phenyl ring of Phe180 is on the surface of the cleft, which might interact with the substituents of xylan, although this residue is not conserved.

The site of cleavage is probably located between subsites B and C or between subsites C and D. Tenkanen *et al.* (1992) have studied the distribution of the produced xylo-oligosaccharides of different substituted and unsubstituted xylans in *T.reesei* xylanase-catalyzed reactions. The hydrolysis of unsubstituted xylan was quite slow, and small amounts of xylotriose and longer xylo-oligomers were produced. The hydrolysis of substituted xylan was more efficient, producing larger amounts of xylo-oligomers. The main product was xylotriose. These data would suggest that the site of cleavage is between subsites C and D.

The reaction mechanism of family G xylanases has been studied experimentally by Gebler *et al.* (1992). This study has revealed that this family of enzymes are 'retaining'. By using the description of Sinnott (1990), the mechanism is e — e. This would mean that the leaving group and the reaction product are both in the equatorial position. Thus, the mechanism would belong to the same class as the widely studied and discussed reaction mechanism of the hen egg white (HEW) lysozyme. During the last few years, many variations for the reaction mechanism of HEW lysozyme have been suggested (Kirby, 1987). As in the HEW lysozyme, the catalytic residues of XYNII have carboxylic acid side chains, which are situated on both sides of the substrate. The distance between the carboxyl groups of catalytic residues is 7.2 Å in the HEW lysozyme. The value has been obtained from the lysozyme structure refined in our laboratory (unpublished data). The corresponding distances identified in this study are 10.7 Å for 'pH 5.0' XYNII and 8.1 Å for 'pH 6.5' XYNII. In XYNII, a new and striking feature is the observed conformational change in the catalytic residue Glu177. We could observe this change through a rising pH, which led to a loss of an acidic proton in Glu177. We suggest that this conformational change also occurs in catalysis (at pH 5.0). The conformational change is able to alter the pK_a value of the side chain of Glu177, making it able to accept or donate a proton, depending on

maps were skeletonized with the Bones program and the result was studied with the program O (Jones *et al.*, 1991). The first Bones maps were calculated using one platinum data set combined from two different crystals and two separate data sets from different crystals soaked in mercury acetate. Two maps were calculated, corresponding to both hands. One map was better, showing the overall shapes of four molecules in the unit cell. However, it was not possible to trace the chain completely from this map. Later, a new map was calculated with the data set from the crystal which was soaked in thioxylose solution and then in methylmercury acetate. This map showed improved continuity and we were able to determine the fold of the protein and assign the sequence. The atomic model for molecule B was built with the O program. This molecule was then shifted visually by using the Bones map to the approximate position of molecule A. The position of the model was refined by using the real-space refinement feature of the O program, and the co-ordinates for both molecules A and B were obtained. Later, the electron density maps were calculated to all heavy-atom derivative data sets using model phases. The electron density maps showed clearly that His144 was the principal binding site in all mercury soaks. However, no density was found for thioxylose. The reason why mercury binding was obtained only in the presence of thioxylose remains unknown.

Structure refinement

The structure refinement was performed with the X-PLOR program (Brünger *et al.*, 1987) with the aid of a graphical user interface developed in our laboratory. The initial model was refined by simulated annealing. Data in the range of 8.0–2.5 Å were used in the first refinement. The standard slow-cooling protocol was used in the refinement, starting at 3000 K and performing 50 steps of molecular dynamics (time step 0.5 fs) and reducing the temperature by 25 K until 300 K was reached. In the first cycle, an *R*-factor was decreased from 59.1 to 37.7% (*R*-factor = $\Sigma(|F_o - F_c|)/\Sigma(F_o)$. The resulting model was studied with the O program, and manual corrections to the structure were made. The refinement cycles, later only with energy minimization, and manual rebuilding were repeated until the *R*-factor decreased to 18.3%. Water molecules were added to the model when the *R*-factor was 24.4%.

The second XYNII model was determined by using the data set based on xylose soaked crystal (pH 6.5). The final model of the native XYNII (pH 5.0) was used as an initial model for the refinement. After some corrections in the side chain and water positions, the final model was obtained. The pictures in this paper have been created by the O (Jones *et al.*, 1991), XtalView (McRee, 1992) and MolScript (Kraulis, 1991) programs. The co-ordinates will be deposited in the Brookhaven Protein Data Bank.

Acknowledgements

We thank Reetta Kallio for her skillful technical assistance. Financial support from the Academy of Finland is gratefully acknowledged.

References

Arase,A., Yomo,T., Urabe,I., Hata,Y., Katsube,Y. and Okada,H. (1993) *FEBS Lett.*, 2, 123–127.

Biely,P. (1985) *Trends Biotechnol.*, 3, 286–290.

Brünger,A.T., Kuriyan,J. and Karplus,M. (1987) *Science*, 235, 458–460.

Davies,G.J., Dodson,G.G., Hubbard,R.E., Tolley,S.P., Dauter,Z., Wilson,K.S., Hjort,C., Mikkelsen,J.M., Rasmussen,G. and Schülein,M. (1993) *Nature*, 365, 362–364.

Fukusaki,E., Panbangred,W., Shinmyo,A. and Okada,H. (1984) *FEBS Lett.*, 171, 197–201.

Gebler,J., Gilkes,N.R., Claeyssens,M., Wilson,D.B., Béguin,P., Wakarchuk,W.W., Kilburn,D.G., Miller,R.C.,Jr, Warren,R.A.J. and Withers,S.G. (1992) *J. Biol. Chem.*, 267, 12559–12561.

Gilbert,H.J. and Hazlewood,G.P. (1993) *J. Gen. Microbiol.*, 39, 187–194.

Gilkes,N.R., Henrissat,B., Kilburn,D.G., Miller,R.C. and Warren,R.A.J. (1991) *Microbiol. Rev.*, 55, 303–315.

Golubev,A.M., Kilimnik,A.Y., Neustroev,K.N. and Pickersgill,R.W. (1993) *J. Mol. Biol.*, 230, 661–663.

Henrissat,B. (1991) *Biochem. J.*, 280, 309–316.

Ito,K., Iwashita,K. and Iwano,K. (1992) *Biosci. Biotech. Biochem.*, 56, 1338–1340.

Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) *Acta Crystallogr.*, A47, 110–119.

Juy,M., Amit,A.G., Alzari,P.M., Poljak R.J., Claeyssens,M., Béguin,P. and Aubert,J.-P. (1992) *Nature*, 357, 89–91.

Keitel,T., Simon,O., Borriss,R. and Heinemann,U. (1993) *Proc. Natl Acad. Sci. USA*, 90, 5287–5291.

Kirby,A.J. (1987) *CRC Crit. Rev. Biochem.*, 22, 283–315.

Ko,E.P., Akatsuka,H., Moriyama,H., Shinmyo,A., Hata,Y., Katsube,Y., Urabe,I. and Okada,H. (1992) *Biochem. J.*, 288, 117–121.

Kraulis,P.J. (1991) *J. Appl. Crystallogr.*, 24, 946–950.

Kubicek,C.P., Messner,R., Gruber,F., Mach,R.L. and Kubicek-Pranz,E.M. (1993) *Enzyme Microbiol. Technol.*, 15, 90–99.

McRee,D.E. (1992) *J. Mol. Graph.*, 10, 44–46.

Moriyama,H., Hata,Y., Yamaguchi,H., Sato,M., Shinmyo,A., Tanaka,N., Okada,H. and Katsube,Y. (1987) *J. Mol. Biol.*, 193, 237–238.

Pickersgill,R.W., Debeire,P., Debeire-Gosselin,M. and Jenkins,J.A. (1993) *J. Mol. Biol.*, 230, 664–666.

Poutanen,K., Rättö,M., Puls,J., and Viikari,L. (1987) *J. Biotechnol.*, 6, 49–60.

Rose,D.R., Birnbaum,G.I., Tan,L.U.L. and Saddler,J.N. (1987) *J. Mol. Biol.*, 194, 755–756.

Rouvinen,J., Bergfors,T., Teeri,T., Knowles,J.K.C. and Jones,T.A. (1990) *Science*, 249, 380–386.

Ryazanova,L.P., Uk,K.D., Beletskaya,O.P. and Kulaev,I.S. (1993) *Biochemistry (USSR)*, 58, 346–352.

Shareck,F., Roy,C., Yaguchi,M., Morosoli,R. and Kluepfel,D. (1991) *Gene*, 107, 75–82.

Sinnott,M.L. (1990) *Chem. Rev.*, 90, 1171–1202.

Steigemann,W. (1992) *Protein, User's Guide*. München.

Tenkanen,M., Puls,J. and Poutanen,K. (1992) *Enzyme Microb. Technol.*, 14, 566–574.

Törrönen,A., Mach,R.L., Messner,R., Gonzalez,R., Kalkkinen,N., Harkki,A. and Kubicek,C.P. (1992) *Bio/Technology*, 10, 1461–1465.

Törrönen,A., Kubicek,C.P. and Henrissat,B. (1993a) *FEBS Lett.*, 321, 135–139.

Törrönen,A., Rouvinen,J., Ahlgren,M., Harkki,A. and Visuri,K. (1993b) *J. Mol. Biol.*, 233, 313–316.

Viikari,L., Kantelinen,A., Poutanen,K. and Ranua,M. (1990) In Kirk,T.K. and Chang,H.M. (eds), *Biotechnology in Pulp and Paper Manufacture. Applications and Fundamental Investigations*. Butterworth–Heinemann, Boston, MA, pp. 145–151.

Viswamitra,M.A., Bhanumoorthy,P., Ramakumar,S., Manjula,M.V., Vithayathil,P.J., Murtjy,S.K. and Naren,A.P. (1993) *J. Mol. Biol.*, 232, 987–988.

Voointholt,R., Kosters,M.T., Vegter,G., Vriend,G. and Hol,W.G.J. (1989) *J. Mol. Graph.*, 7, 243–245.

Vyas,N.K. (1991) *Curr. Opin. Struct. Biol.*, 1, 732–740.

Wong,K.K.Y. and Saddler,J.N. (1992) In Visser,J. (ed.), *Xylan and Xylanases*. Elsevier, Amsterdam, pp. 171–187.

Yang,R.C.A., MacKenzie,C.R. and Narang,S.A. (1988) *Nucleic Acids Res.*, 16, 7187.

Zappe,H., Jones,W.A. and Woods,D.R. (1990) *Nucleic Acids Res.*, 18, 2179.

Received on February 14, 1994; revised on March 22, 1994

Noted added in proof

Recently, Campbell and coworkers have solved the structures of two xylanases from family G, namely xylanases from *Trichoderma harzianum* and *Bacillus circulans* [Campbell,R.L., Rose,D.R., Wakarchuk,W.W., To,R., Sung,W. and Yaguchi,M. (1993) In Suominen,P. and Reinikainen,T. (eds), *Proceedings of the Second TRICEL Symposium on Trichoderma reesei Cellulases and Other Hydrolases*. Foundations for Biotechnical and Industrial Fermentation Research 8, Espoo, pp. 63–72].

New Atty Docket No.: GOWL-018/00US/306506-2025 PATENT
Former Atty Docket No.: 07121-003U1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Wing L. SUNG	Confirmation No.: 2196
Application No.: 09/990,874	Examiner: M. N. RAO
Filed: 11/21/2001	Art Unit: 1652

FOR: XYLANASES WITH ENHANCED THERMOPHILICITY AND LKALOPHILICITY

Commissioner for Patents
U.S. Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

REVOCATION AND NEW POWER BY ASSIGNEE AND STATEMENT UNDER 37 C.F.R. §3.73(b)

Dear Sir:

The Assignee of the entire right, title, and interest in the above-identified application hereby revokes all previously granted powers and grants the registered practitioners of Cooley Godward Kronish LLP included in the Customer Number provided below power to act, prosecute, and transact all business in the U.S. Patent and Trademark Office in connection with this application, any applications claiming priority to this application, and any patents issuing therefrom.

The assignee certifies that to the best of its knowledge and belief it is the owner of the entire right, title, and interest in and to the above-identified application as evidenced by:

[] An assignment document, a copy of which is enclosed herewith;

[x] An assignment previously recorded in the U.S. Patent and Trademark Office at Reel/Frame No. 012657/0447.

Please direct all telephone calls and correspondence to:

Edward I. Amaya
COOLEY GODWARD KRONISH LLP
1200 19th Street, N.W., 5th Floor
Washington, DC 20036
Tel: 202-842-7800
Fax: 202-842-7899

CUSTOMER NUMBER: **58249**

The undersigned (whose title is supplied below) is empowered to sign this statement on behalf of the assignee.

Date: 15 June 2007

Signature: Marielle Piché
Name: Marielle Piché
Title: Secretary General
Company: National Research Council